

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 January 2006 (19.01.2006)

PCT

(10) International Publication Number  
**WO 2006/007400 A2**

(51) International Patent Classification:  
C12N 1/00 (2006.01)

(21) International Application Number:

PCT/US2005/021297

(22) International Filing Date: 15 June 2005 (15.06.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/580,448 16 June 2004 (16.06.2004) US

(71) Applicant (for all designated States except US):  
METABOLEX, INC. [US/US]; 3876 Bay Center  
Place, Hayward, CA 94545 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): MOODIE, Shonna  
A. [GB/US]; 2091 Golden Gate, San Francisco, CA  
94111 (US). ZHANG, Fang [CN/US]; 566 Simon Street,  
Hayward, CA 94541 (US). RACK, Paul G. [US/US];  
34301 Parker Court, Fremont, CA 94555 (US). SHANG,  
Jin [US/US]; 3142 Flowers Lane, Palo Alto, CA 94306  
(US). LAVAN, Brian E. [GB/US]; 2368 21st Avenue, San  
Francisco, CA 94122 (US). ALLAN, Bernard [IE/US];  
940 Guerrero Street, San Francisco, CA 94110 (US).  
WONG, Chi-Wai [CN/US]; 28073 Thorup Lane, Hay-  
ward, CA 94542 (US). GREGOIRE, Francine [BE/US];  
1044 Carol Lane, Lafayette, CA 94549 (US). PEREZ,  
Grace [US/US]; 6115 Thornton Avenue, Newark, CA  
94560 (US). WATERS, Steve [US/US]; 1 Lobelia Lane,  
San Ramon, CA 94583 (US).

(74) Agents: HINSCH, Matthew, E. et al.; Townsend and  
Townsend and Crew LLP, Two Embarcadero Center, 8th  
Floor, San Francisco, CA 94111 (US).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AI., AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GH, GI, GM, GR, GU, HD, IL, IN, IS, JP, KE,  
KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,  
MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ,  
OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL,  
SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,  
VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,  
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO,  
SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,  
GQ, GW, ML, MR, NE, SN, TD, TG).

**Declaration under Rule 4.17:**

— of inventorship (Rule 4.17(iv)) for US only

**Published:**

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: METHODS OF DIAGNOSING AND TREATING OBESITY, DIABETES AND INSULIN RESISTANCE

(57) Abstract: The present invention provides compositions and methods for diagnosing and treating obesity, diabetes and insulin resistance. In particular, the invention provides methods of identifying modulators of the polynucleotides or polypeptides of the invention and using those modulators to treat obesity and/or diabetes, as well as methods of diagnosing obesity and/or diabetes by measuring the levels of the polynucleotides or polypeptides of the invention in a patient.

WO 2006/007400 A2

## **Methods of Diagnosing & Treating Obesity, Diabetes and Insulin Resistance**

### **CROSS-REFERENCE TO RELATED PATENT APPLICATIONS**

5           [01]    The present application claims priority to U.S. Provisional Patent Application No. 60/ 580,448, filed June 16, 2004, which is incorporated by reference in its entirety for all purposes.

### **BACKGROUND OF THE INVENTION**

10           [02]    Obesity has reached epidemic proportions globally with more than 1 billion adults overweight- at least 300 million of them clinical obese- and is a major contributor to the global burden of chronic disease and disability. Overweight and obesity leads to adverse metabolic effect on blood pressure, cholesterol, triglycerides and insulin resistance. The non-fatal but debilitating health problems associated with obesity include  
15   respiratory difficulties, chronic musculoskeletal problems, skin problems and infertility. The more life-threatening problems fall into four main areas: cardiovascular disease problems, conditions associated with insulin resistance such as Type 2 diabetes, certain types of cancers especially the hormonally related and large-bowel cancers, and gall bladder disease. The likelihood of developing Type 2 diabetes and hypertension rises steeply with increasing body  
20   fatness. Weight reduction leads to correction of a number of obesity- associated endocrine and metabolic disorders.

          [03]    Effective weight management for individuals and groups at risk of developing obesity involves a range of long term strategies. These include prevention, weight maintenance, management of co-morbidities and weight loss. Existing treatment strategies  
25   include calorific restriction programs, surgery (gastric stapling) and drug intervention. The currently available anti-obesity drugs can be divided into two classes: central acting and peripheral acting. Three marketed drugs are Xenical (Orlistat), Merida (Sibutramine) and Adipex-P (Phentermine). Xenical is a non-systemic acting GI lipase inhibitor which is indicated for short and long term obesity management. Merida reduces food intake by re-  
30   uptake inhibition of primarily norepinephrine and serotonin. Adipex-P is a phenteramine with sympathomimetic activities and suppresses appetite. It is indicated only for short term

use. A more drastic solution to permanent weight loss is surgery and a gastric by-pass which limits absorption of calories through massive reduction in stomach size.

[04] Carrying extra body weight and body fat go hand and hand with the development of diabetes. People who are overweight (BMI greater than 25) are at a much greater risk of developing type 2 diabetes than normal weight individuals. Almost 90% of people with type 2 diabetes are overweight.

[05] Diabetes mellitus can be divided into two clinical syndromes, Type 1 and Type 2 diabetes mellitus. Type 1, or insulin-dependent diabetes mellitus (IDDM), is a chronic autoimmune disease characterized by the extensive loss of beta cells in the pancreatic Islets of Langerhans, which produce insulin. As these cells are progressively destroyed, the amount of secreted insulin decreases, eventually leading to hyperglycemia (abnormally high level of glucose in the blood) when the amount of secreted insulin drops below the level required for euglycemia (normal blood glucose level). Although the exact trigger for this immune response is not known, patients with IDDM have high levels of antibodies against proteins expressed in pancreatic beta cells. However, not all patients with high levels of these antibodies develop IDDM.

[06] Type 2 diabetes (also referred to as non-insulin dependent diabetes mellitus (NIDDM)) develops when muscle, fat and liver cells fail to respond normally to insulin. This failure to respond (called insulin resistance) may be due to reduced numbers of insulin receptors on these cells, or a dysfunction of signaling pathways within the cells, or both. The beta cells initially compensate for this insulin resistance by increasing insulin output. Over time, these cells become unable to produce enough insulin to maintain normal glucose levels, indicating progression to Type 2 diabetes.

[07] Type 2 diabetes is brought on by a combination of genetic and acquired risk factors - including a high-fat diet, lack of exercise, and aging. Worldwide, Type 2 diabetes has become an epidemic, driven by increases in obesity and a sedentary lifestyle, widespread adoption of western dietary habits, and the general aging of the population in many countries. In 1985, an estimated 30 million people worldwide had diabetes -- by 2000, this figure had increased 5-fold, to an estimated 154 million people. The number of people with diabetes is expected to double between now and 2025, to about 300 million.

[08] Type 2 diabetes is a complex disease characterized by defects in glucose and lipid metabolism. Typically there are perturbations in many metabolic parameters including increases in fasting plasma glucose levels, free fatty acid levels and

triglyceride levels, as well as a decrease in the ratio of HDL/LDL. As discussed above, one of the principal underlying causes of diabetes is thought to be an increase in insulin resistance in peripheral tissues, principally muscle and fat.

[09] Therapies aimed at reducing peripheral insulin resistance are available.

5 The most relevant to this invention are drugs of the thiazolidinedione (TZD) class namely troglitazone, pioglitazone, and rosiglitazone. In the US these have been marketed under the names Rezulin™, Avandia™ and Actos™, respectively. The principal effect of these drugs is to improve glucose homeostasis. Notably in diabetics treated with TZDs there are increases in peripheral glucose disposal rates indicative of increased insulin sensitivity in both muscle  
10 and fat.

[10] The molecular target of TZDs is a member of the PPAR family of ligand-activated transcription factors called PPAR gamma. This transcription factor is highly expressed in adipose tissue with much lower levels being observed in muscle. Binding of TZDs to PPAR gamma in target cells and tissues such as fat and muscle brings about a  
15 change in gene expression. The link between TZD-altered gene expression in fat and muscle and increased insulin sensitivity is unknown. The present invention addresses this and other problems.

#### BRIEF SUMMARY OF THE INVENTION

20 The present invention provides methods for identifying an agent for treating an obese, diabetic or pre-diabetic individual. In some embodiments, the method comprising the steps of: (i) contacting an agent to a polypeptide encoded by a polynucleotide that is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a  
25 wash in 0.2X SSC, and 0.1% SDS at 55°C, wherein the polypeptide optionally has the activity listed in Table 1; and (ii) selecting an agent that modulates the expression or activity of the polypeptide or that binds to the polypeptide, thereby identifying an agent for treating an obese, diabetic or pre-diabetic individual.

**Table 1: List of Polypeptides, SEQ ID numbers and Proposed Activity**

PROTEIN NAME	SEQ ID NO:	PROPOSED ACTIVITY (S)
Adlican	2, 4, 6	Signal Transduction
ALDH1A3	8, 10, 12	Aldehyde dehydrogenase
ALK7	14, 16, 18	Receptor Serine/threonine protein kinase



C3AR1	20, 22, 24	G-protein coupled receptor for Complement Component C3a
CALCRL	26, 28, 30	G-protein coupled receptor for adrenomedullin or calcitonin gene-related protein
CCL13	32, 33, 35, 37	Chemokine
CCL8	39, 40, 42, 44	Chemokine
CH13L1	46, 47, 49, 51	Glycosyl hydrolase
CR1	53, 55	Transmembrane receptor for Complement Component C3b-C4b
CSFR1	57, 59, 61	Receptor tyrosine kinase
CTSK	63, 64, 66, 68	Cysteine protease
CXCR4	70, 72, 74	G-protein coupled receptor for CXCL12
DDAH2	76, 78, 80	Amidinotransferase
DERP7	82, 84, 86	7 transmembrane protein
ENDOGLYX1	88, 90, 92	Modifier of extracellular matrix
ETL	94, 96, 98	G-protein coupled receptor
FLJ12389	100, 102, 104, 106, 108	Acetoacetate CoA ligase
FZD4	110, 112, 114, 116	7 Transmembrane receptor for Wnt proteins
GLIPR1	118, 120, 122	Apoptosis Regulator
GPR105	124, 126, 128	G-protein coupled receptor for UDP sugars
GPR146	130, 132, 134	G-protein coupled receptor
GPR30	136, 138, 140	G-protein coupled receptor
GPR65	142, 144, 146	G-protein coupled receptor for psychosine
HTR2B	148, 150, 152	G-protein coupled receptor for serotonin
ITGB2	154, 156, 158	Cell Adhesion
ITIH5	160, 161, 163	Extracellular matrix stabilization
LGALS12	165, 167, 169, 171, 173, 175, 177, 179	Apoptosis Regulator
NMB	181, 182, 184, 186	Ligand for the G-protein coupled receptor, NMBR
NNAT	188, 190, 192, 194	Regulator of ion channels
OLFM2	196, 197, 199, 201	Secreted protein
OPN3	203, 205, 207, 209, 211	G-protein coupled receptor

PTPRE	213, 215, 217, 219, 221	Protein tyrosine phosphatase
RDC1	223, 225, 227	G-protein coupled receptor
SLIT2	229, 230, 232, 234	Ligand for roundabout receptor, ROBO1
TNFRSF21	236, 238, 240	Transmembrane receptor
TNFSF13B	242, 243, 245, 247, 249	Ligand for TNFRSF13B, TNFRSF13C and TNFRSF17B
TNFSF14	251, 252, 254, 256, 258, 260	Ligand for TNFRSF14
TPSB2	262, 263, 265, 267	Serine protease
WISP2	269, 270, 272, 274	Growth Regulator

In some embodiments, the polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272, 274 or a protein domain thereof. In some embodiments, the method further comprises detecting whether the selected agent modulates weight and/or obesity. In some embodiments, the method further comprising detecting whether the selected agent modulates insulin sensitivity.

[11] In some embodiments, step (ii) comprises selecting an agent that modulates expression of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that modulates the activity of the polypeptide. In some embodiments, step (ii) comprises selecting an agent that specifically binds to the polypeptide.

[12] In some embodiments, the polypeptide is expressed in a cell and the cell is contacted with the agent. In some embodiments, the polypeptide comprises SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171,

173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.

[13] The present invention also provides methods of reducing body weight  
5 in an animal. In some embodiments, the methods comprise administering to the animal an effective amount of an agent that modulates the activity or expression of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138,  
10 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.

[14] In some embodiments, the agent is selected by a method comprising (i)  
15 contacting an agent to a mixture comprising a polypeptide encoded by a polynucleotide is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C, wherein the polypeptide optionally has the activity listed in Table 1; and (ii) selecting an agent that modulates the expression or activity  
20 of the polypeptide or that binds to the polypeptide.

[15] In some embodiments, the agent is an antibody. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the animal is a human.

[16] The present invention also provides methods of treating a diabetic or pre-diabetic animal. In some embodiments, the method comprising administering to the  
25 animal a therapeutically effective amount of an agent that modulates the activity or expression of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161,  
30 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274. In some embodiments, the polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32,

33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272, 274 or a protein domain thereof.

[17] In some embodiments, the agent is selected by a method comprising (i) contacting an agent to a mixture comprising a polypeptide encoded by a polynucleotide that hybridizes to a nucleic acid encoding SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274 in 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C; and (ii) selecting an agent that modulates the expression or activity of the polypeptide or that binds to the polypeptide.

[18] In some embodiments, the agent is an antibody. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the animal is a human.

[19] The present invention also provides methods of introducing an expression cassette into a cell. In some embodiments, the methods comprise introducing into the cell an expression cassette comprising a promoter operably linked to a polynucleotide encoding a polypeptide, wherein the polynucleotide is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C, and the polypeptide optionally has the activity listed in Table 1. In some embodiments, the polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203,

205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272, 274 or a protein domain thereof.

[20] In some embodiments, the polypeptide comprises SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.

[21] In some embodiments, the cell is selected from the group consisting of an adipocyte and a skeletal muscle cell.

[22] In some embodiments, the method further comprises introducing the cell into a human. In some embodiments, the human is obese. In some embodiments, the human is diabetic. In some embodiments, the human is prediabetic. In some embodiments, the cell is from the human.

[23] The present invention also provides methods of diagnosing an individual who has obesity, Type 2 diabetes or has a predisposition for diabetes or obesity. In some embodiments, the method comprises detecting in a sample from the individual the level of a polypeptide or the level of a polynucleotide encoding the polypeptide, wherein the polynucleotide is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C, wherein a modulated level of the polypeptide or polynucleotide in the sample compared to a level of the polypeptide or polynucleotide in either a lean individual or a previous sample from the individual indicates that the individual is obese or diabetic or has a predisposition for diabetes or obesity.

[24] In some embodiments, the detecting step comprises contacting the sample with an antibody that specifically binds to the polypeptide. In some embodiments, the amino acid sequence comprises SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156,

158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274. In some embodiments, the detecting step comprises

- 5 quantifying mRNA encoding the polypeptide. In some embodiments, the mRNA is reverse transcribed and amplified in a polymerase chain reaction. In some embodiments, the sample is a blood, urine or tissue sample.

[25] The present invention provides for an isolated nucleic acid that is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C. In some embodiments, the polynucleotide comprises SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 34, 36, 38, 41, 43, 45, 48, 50, 52, 54, 56, 58, 60, 62, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 183, 185, 187, 189, 191, 193, 195, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 231, 233, 235, 237, 239, 241, 244, 246, 248, 250, 253, 255, 257, 259, 261, 264, 266, 268, 271 or 273. In some embodiments, the polynucleotide encodes SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.

[26] The present invention also provides expression cassettes comprising a heterologous promoter operably linked to a nucleic acid that is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at 42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C.

[27] The present invention also provides host cells transfected with nucleic acids that is substantially identical to or hybridizes to a nucleic acid encoding a polypeptide listed in Table 1 under hybridization conditions of 50% formamide, 5X SSC, and 1% SDS at

42°C followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C. In some embodiments, the host cell is a human cell. In some embodiments, the host cell is a bacterium.

[28] The present invention also provides isolated polypeptides comprising an amino acid sequence at least 70% identical to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272, 274 or fragments thereof. In some embodiments, the polypeptide comprises SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.

[29] The present invention also provides antibodies that specifically bind to a polypeptide selected from the groups consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.

[30] The present invention also provides pharmaceutical compositions comprising polypeptides comprising an amino acid sequence at least 70% identical to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169,

171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272, 274 or fragments thereof, and a pharmaceutically-acceptable excipient.

5

## DEFINITIONS

[31] "Insulin sensitivity" refers to the ability of a cell or tissue to respond to insulin. Responses include, e.g., glucose uptake of a cell or tissue in response to insulin stimulation. Sensitivity can be determined at an organismal, tissue or cellular level. For example, blood or urine glucose levels following a glucose tolerance test are indicative of insulin sensitivity. Other methods of measuring insulin sensitivity include, e.g., measuring glucose uptake (*see, e.g.,* Garcia de Herreros, A., and Birnbaum, M. J. *J. Biol. Chem.* 264, 19994-19999 (1989); Klip, A., Li, G., and Logan, W.J. *Am. J. Physiol.* 247, E291-296 (1984)), measuring the glucose infusion rate (GINF) into tissue such as the skeletal muscle (*see, e.g.,* Ludvik *et al., J. Clin. Invest.* 100:2354 (1997); Frias *et al., Diabetes Care* 23:64, (2000)) and measuring sensitivity of GLUT4 translocation (e.g., as described herein) in response to insulin.

[32] As used herein, an overweight person has a body mass index (BMI)  $\geq$  25 and an "obese" person has a BMI  $\geq$ 30. BMI is calculated as the weight in kilograms divided by the square of the height in meters.

[33] The term "waist-to-hip ratio or WHR" is the ratio of a person's waist circumference to hip circumference. For most people, carrying extra weight around their middle increases health risks more than carrying extra weight around their hips or thighs. For both men and women, a waist-to-hip ratio of 1.0 or higher is considered "at risk" or in the danger zone for undesirable health consequences, such as heart disease and other ailments connected with being overweight.

[34] The term "adipogenic," when used in reference to cells refers to a cell which can become an adipocyte. An "adipogenic factor" refers to a factor (including, e.g., a protein (or glycoprotein)) that can induce or stimulate the differentiation of cells into an adipocyte. Exemplary adipogenic factors include, e.g., \_Wnt10b, Pref-1, ADF and TNF-alpha.

[35] The term "lipid metabolism" refers to the in vivo process of catabolism (decomposition) and anabolism (accumulation) of lipids (e.g., triglycerides derived from food) and is intended to include, in the broad sense, reactions for transforming lipids into



energy, biosynthesis of fatty acids, acylglycerol, phospholipid metabolism and cholesterol metabolism.

[36] "Activity" of a polypeptide of the invention refers to structural, regulatory, or biochemical functions of a polypeptide in its native cell or tissue. Examples of activity of a polypeptide include both direct activities and indirect activities. Exemplary direct activities are the result of direct interaction with the polypeptide, , e.g., enzymatic activity, ligand binding, production or depletion of second messengers (e.g., cAMP, cGMP, IP<sub>3</sub>, DAG, or Ca<sup>2+</sup>), ion flux, phosphorylation levels, transcription levels, and the like. Exemplary indirect activities are observed as a change in phenotype or response in a cell or tissue to a polypeptide's directed activity, e.g., loss of body weight or molecular events associated with loss of body weight or obesity or modulating insulin sensitivity of a cell as a result of the interaction of the polypeptide with other cellular or tissue components.

[37] "Predisposition for diabetes" occurs in a person when the person is at high risk for developing diabetes. A number of risk factors are known to those of skill in the art and include: genetic factors (e.g., carrying alleles that result in a higher occurrence of diabetes than in the average population or having parents or siblings with diabetes); overweight (e.g., body mass index (BMI) greater or equal to 25 kg/m<sup>2</sup>); habitual physical inactivity, race/ethnicity (e.g., African-American, Hispanic-American, Native Americans, Asian-Americans, Pacific Islanders); previously identified impaired fasting glucose or impaired glucose tolerance, hypertension (e.g., greater or equal to 140/90 mmHg in adults); HDL cholesterol less than or equal to 35 mg/dl; triglyceride levels greater or equal to 250 mg/dl; a history of gestational diabetes or delivery of a baby over nine pounds; and/or polycystic ovary syndrome. See, e.g., "Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus" and "Screening for Diabetes" *Diabetes Care* 25(1): S5-S24 (2002).

[38] A "lean individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level less than 100 mg/dl or a 2 hour PG reading of 140 mg/dl. "Fasting" refers to no caloric intake for at least 8 hours. A "2 hour PG" refers to the level of blood glucose after challenging a patient to a glucose load containing the equivalent of 75g anhydrous glucose dissolved in water. The overall test is generally referred to as an oral glucose tolerance test (OGTT). See, e.g., *Diabetes Care*, 2003, 26(11): 3160-3167 (2003). The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean

individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[39] A "pre-diabetic individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 100 mg/dl but less than 126 mg/dl or a 2 hour PG reading of greater than 140 mg/dl but less than 200mg/dl. A  
5 "diabetic individual," when used to compare with a sample from a patient, refers to an adult with a fasting blood glucose level greater than 126 mg/dl or a 2 hour PG reading of greater than 200 mg/dl.

[40] An "agonist" refers to an agent that binds to, stimulates, increases,  
10 activates, facilitates, enhances activation, sensitizes or up regulates the activity or expression of a polypeptide of the invention.

[41] An "antagonist" refers to an agent that binds to, partially or totally blocks stimulation, decreases, prevents, delays activation, inactivates, desensitizes, or down regulates the activity or expression of a polypeptide of the invention.

[42] "Antibody" refers to a polypeptide substantially encoded by an  
15 immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda.  
20 Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

[43] An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus  
25 of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain ( $V_L$ ) and variable heavy chain ( $V_H$ ) refer to these light and heavy chains respectively.

[44] Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for  
30 example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce  $F(ab)_2$ , a dimer of Fab which itself is a light chain joined to  $V_H-C_H1$  by a disulfide bond. The  $F(ab)_2$  may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the  $F(ab)_2$  dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (see, Paul (Ed.) *Fundamental Immunology*,

Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv).

[45] The terms "peptidomimetic" and "mimetic" refer to a synthetic chemical compound that has substantially the same structural and functional characteristics of the antagonists or agonists of the invention. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed "peptide mimetics" or "peptidomimetics" (Fauchere, J. *Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS* p. 392 (1985); and Evans *et al. J. Med. Chem.* 30:1229 (1987), which are incorporated herein by reference). Peptide mimetics that are structurally similar to therapeutically useful peptides may be used to produce an equivalent or enhanced therapeutic or prophylactic effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (i.e., a polypeptide that has a biological or pharmacological activity), such as a polypeptide exemplified in this application, but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of, e.g., -CH<sub>2</sub>NH-, -CH<sub>2</sub>S-, -CH<sub>2</sub>-CH<sub>2</sub>-, -CH=CH- (cis and trans), -COCH<sub>2</sub>-, -CH(OH)CH<sub>2</sub>-, and -CH<sub>2</sub>SO-. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity. For example, a mimetic composition is within the scope of the invention if it is capable of carrying out the binding or other activities of an agonist or antagonist of a polypeptide of the invention.

[46] The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[47] The term "isolated," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It may be in a homogeneous state although it can be in either a

dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames that flank the gene and encode a protein other than the gene of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

[48] The term "nucleic acid" or "polynucleotide" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.* (1992); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

[49] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds.

[50] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline,  $\gamma$ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to

compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an  $\alpha$  carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but  
5 retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but which functions in a manner similar to a naturally occurring amino acid.

[51] Amino acids may be referred to herein by either the commonly known  
10 three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[52] "Conservatively modified variants" applies to both amino acid and  
nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively  
15 modified variants" refers to those nucleic acids that encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the  
codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every  
20 position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic  
25 acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

[53] As to amino acid sequences, one of skill will recognize that individual  
30 substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in

the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[54] The following eight groups each contain amino acids that are conservative substitutions for one another:

- 5 1) Alanine (A), Glycine (G);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 10 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and
- 8) Cysteine (C), Methionine (M)

(see, e.g., Creighton, *Proteins* (1984)).

[55] "Percentage of sequence identity" is determined by comparing two  
15 optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (e.g., a polypeptide of the invention), which does not comprise additions or deletions, for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic  
20 acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[56] The terms "identical" or percent "identity," in the context of two or  
25 more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same sequences. Sequences are "substantially identical" if two sequences have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum  
30 correspondence over a comparison window, designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection, or across the entire sequence where not indicated. The invention provides polypeptides or polynucleotides that are substantially identical to the polypeptides or polynucleotides, respectively, exemplified herein (e.g., SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,

15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39,  
40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64,  
65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89,  
90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111,  
5 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130,  
131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149,  
150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168,  
169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187,  
188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 201, 202, 203, 204, 205, 206, 207,  
10 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226,  
227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245,  
246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264,  
265, 266, 267, 268, 269, 270, 271, 272, 273 or 274). This definition also refers to the  
complement of a test sequence. Optionally, the identity exists over a region that is at least  
15 about 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or  
more nucleotides in length.

[57] For sequence comparison, typically one sequence acts as a reference  
sequence, to which test sequences are compared. When using a sequence comparison  
algorithm, test and reference sequences are entered into a computer, subsequence coordinates  
20 are designated, if necessary, and sequence algorithm program parameters are designated.  
Default program parameters can be used, or alternative parameters can be designated. The  
sequence comparison algorithm then calculates the percent sequence identities for the test  
sequences relative to the reference sequence, based on the program parameters.

[58] A "comparison window", as used herein, includes reference to a  
25 segment of any one of the number of contiguous positions selected from the group consisting  
of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in  
which a sequence may be compared to a reference sequence of the same number of  
contiguous positions after the two sequences are optimally aligned. Methods of alignment of  
sequences for comparison are well known in the art. Optimal alignment of sequences for  
30 comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman  
(1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and  
Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and  
Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of  
these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics

Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Ausubel et al., Current Protocols in Molecular Biology* (1995 supplement)).

[59] Two examples of algorithms that are suitable for determining percent  
5 sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which  
are described in Altschul *et al.* (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul *et al.*  
(1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is  
publicly available through the National Center for Biotechnology Information  
(<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring  
10 sequence pairs (HSPs) by identifying short words of length W in the query sequence, which  
either match or satisfy some positive-valued threshold score T when aligned with a word of  
the same length in a database sequence. T is referred to as the neighborhood word score  
threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for  
initiating searches to find longer HSPs containing them. The word hits are extended in both  
15 directions along each sequence for as far as the cumulative alignment score can be increased.  
Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward  
score for a pair of matching residues; always > 0) and N (penalty score for mismatching  
residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the  
cumulative score. Extension of the word hits in each direction are halted when: the  
20 cumulative alignment score falls off by the quantity X from its maximum achieved value; the  
cumulative score goes to zero or below, due to the accumulation of one or more negative-  
scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm  
parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN  
program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation  
25 (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the  
BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the  
BLOSUM62 scoring matrix (*see* Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA*  
89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both  
strands.

30 [60] The BLAST algorithm also performs a statistical analysis of the  
similarity between two sequences (*see, e.g., Karlin and Altschul* (1993) *Proc. Natl. Acad. Sci.*  
*USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the  
smallest sum probability (P(N)), which provides an indication of the probability by which a  
match between two nucleotide or amino acid sequences would occur by chance. For example,



a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[61] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[62] The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

[63] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength pH. The  $T_m$  is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at  $T_m$ , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as

formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5X SSC, and 1% SDS, incubating at 42°C, or 5X SSC, 1% SDS, incubating at 65°C, with wash in 0.2X SSC, and  
5 0.1% SDS at 55°C, 60°C, or 65°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

[64] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created  
10 using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is  
15 at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

[65] The phrase "a nucleic acid sequence encoding" refers to a nucleic acid which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid sequence of a specific protein or peptide, or a binding site for a trans-  
20 acting regulatory agent. This phrase specifically encompasses degenerate codons (*i.e.*, different codons which encode a single amino acid) of the native sequence or sequences that may be introduced to conform with codon preference in a specific host cell.

[66] The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has  
25 been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

[67] The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not  
30 found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a

coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

[68] An "expression vector" is a nucleic acid construct, generated  
5 recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

[69] The phrase "specifically (or selectively) binds to an antibody" or  
10 "specifically (or selectively) immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding  
15 to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised against a protein having an amino acid sequence encoded by any of the polynucleotides of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins, except for polymorphic variants. A variety of immunoassay formats may be used to select  
20 antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. *See*, Harlow and Lane *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, NY (1988) for a description of immunoassay formats and conditions that can be used to determine specific  
25 immunoreactivity. Typically, a specific or selective reaction will be at least twice the background signal or noise and more typically more than 10 to 100 times background.

[70] "Inhibitors," "activators," and "modulators" of expression or of  
activity are used to refer to inhibitory, activating, or modulating molecules, respectively, of expression of the polypeptides of the invention as determined using *in vitro* or *in vivo* assays  
30 to monitor expression or activity. Modulators encompass *e.g.*, ligands, agonists, antagonists, their homologs and mimetics, as well as the polypeptides of the invention, or fragments thereof with antagonist activity or that act to increase overall polypeptide activity (*i.e.*, fragments that have at least some of the activity of the full-length protein). In some cases, fragments of the polypeptides of the invention are at least 20, 50, 75 or 100 amino acids in

length. The term "modulator" includes inhibitors and activators. Inhibitors are agents that, *e.g.*, inhibit expression of a polypeptide of the invention or bind to, partially or totally block stimulation, decrease, prevent, delay activation, inactivate, desensitize, or down regulate the activity of a polypeptide of the invention, *e.g.*, antagonists. Activators are agents that, *e.g.*, induce or activate the expression of a polypeptide of the invention or bind to, stimulate, increase, open, activate, facilitate, or enhance activation, sensitize or up regulate the activity of a polypeptide of the invention, *e.g.*, agonists. Modulators include naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Such assays for inhibitors and activators include, *e.g.*, applying putative modulator compounds to cells expressing a polypeptide of the invention and then determining the functional effects on a polypeptide of the invention activity, as described above. Samples or assays comprising a polypeptide of the invention that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of effect. Control samples (untreated with modulators) are assigned a relative activity value of 100%. Inhibition of a polypeptide of the invention is achieved when the polypeptide activity value relative to the control is about 80%, optionally 50% or 25, 10%, 5% or 1%. Activation of the polypeptide is achieved when the polypeptide activity value relative to the control is 110%, optionally 150%, optionally 200, 300%, 400%, 500%, or 1000-3000% or more higher.

## DETAILED DESCRIPTION OF THE INVENTION

### I. INTRODUCTION

[71] The present application demonstrates that, surprisingly, modulated levels of mRNA comprising sequences of the invention occur in human adipose tissue collected from either insulin resistant obese non-diabetics or from type 2 diabetic individuals compared to levels of the mRNA in the lean, non-diabetic individuals. Insulin resistant obese individuals are generally predisposed to become type II diabetics. Therefore, the modulation of the sequences in the study described herein indicates the sequences' involvement in obesity, diabetes and/or pre-diabetes.

[72] Without intending to limit the invention to a particular mechanism of action, it is believed that modulation of the expression or activity of the polypeptides or polynucleotides of the invention is beneficial in treating obesity, diabetic, pre-diabetic or insulin resistant, non-diabetic patients. Furthermore, modulated levels of the polypeptides of the invention are indicative of insulin resistance, obesity, diabetes or a predisposition for

obesity and/or diabetes. Thus, the detection of a polypeptide of the invention is useful for diagnosis of obesity, predisposition for obesity and/or diabetes, diabetes and/or insulin resistance.

[73] This invention also provides methods of using polypeptides of the invention and modulators of the polypeptides of the invention to diagnose and treat obesity, diabetes, pre-diabetes (including insulin resistant individuals) and related metabolic diseases. The present method also provides methods of identifying modulators of expression or activity of the polypeptides of the invention. Such modulators are useful for treating obesity and/or Type 2 diabetes as well as the pathological aspects of obesity (e.g., increased risk for cardiovascular disease, hypertension or cancer) and/or diabetes (e.g., insulin resistance).

## II. GENERAL RECOMBINANT NUCLEIC ACID METHODS FOR USE WITH THE INVENTION

[74] In numerous embodiments of the present invention, nucleic acids encoding a polypeptide of the present invention will be isolated and cloned using recombinant methods. Such embodiments are used, *e.g.*, to isolate polynucleotides identical or substantially identical to SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 34, 36, 38, 41, 43, 45, 48, 50, 52, 54, 56, 58, 60, 62, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 183, 185, 187, 189, 191, 193, 195, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 231, 233, 235, 237, 239, 241, 244, 246, 248, 250, 253, 255, 257, 259, 261, 264, 266, 268, 271 or 273 for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from an polypeptide or polynucleotide of the invention, to monitor gene expression, for the isolation or detection of sequences in different species, for diagnostic purposes in a patient, *e.g.*, to detect mutations in a polypeptide or polynucleotide of the invention or to detect expression levels of nucleic acids or polypeptides. In some embodiments, the sequences encoding the polypeptides of the invention (or polypeptides comprising fragments of the polypeptides of the invention) are operably linked to a heterologous promoter. In some cases, fragments of the polypeptides of the invention are at least 20, 50, 75 or 100 amino acids in length. The polypeptides of the invention can be linked to heterologous amino acid sequences using recombinant DNA technology. In one embodiment, the nucleic acids of the invention are from any mammal, including, in particular, *e.g.*, a human, a mouse, a rat, etc.

[75] Polynucleotides, including expression cassettes, encoding polypeptides of the invention can be introduced into cells and optionally expressed in the cells.

Polynucleotides of the invention can be introduced into eukaryotic or prokaryotic cells, including adipocyte or muscle cells. The cells can be primary cells or cell lines.

5           **A. General Recombinant Nucleic Acid Methods**

[76] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

[77] For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Proteins sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

[78] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter *et. al.*, *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom.* 255:137-149 (1983).

[79] The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace *et al.*, *Gene* 16:21-26 (1981).

**B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding Desired Proteins**

[80] In general, the nucleic acids encoding the subject proteins are cloned from DNA sequence libraries that are made to encode cDNA or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences disclosed herein, which provide a reference for PCR primers and defines suitable regions for isolating probes specific for the polypeptides or polynucleotides of the invention. Alternatively, where the sequence is cloned

into an expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against a polypeptide of interest, including those disclosed herein.

[81] Methods for making and screening genomic and cDNA libraries are well known to those of skill in the art (*see, e.g., Gubler and Hoffman Gene* 25:263-269 (1983); Benton and Davis *Science*, 196:180-182 (1977); and Sambrook, *supra*).

[82] Briefly, to make the cDNA library, one should choose a source that is rich in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue and either mechanically sheared or enzymatically digested to yield fragments of preferably about 5-100 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*, and the recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein *et al., Proc. Natl. Acad. Sci. USA.*, 72:3961-3965 (1975).

[83] An alternative method combines the use of synthetic oligonucleotide primers with polymerase extension on an mRNA or DNA template. Suitable primers can be designed from specific sequences disclosed herein. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be useful, for example, to clone nucleic acids encoding specific proteins and express said proteins, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a polypeptide of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (*see, U.S. Patent Nos. 4,683,195 and 4,683,202*). Genes amplified by a PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

[84] Appropriate primers and probes for identifying the genes encoding a polypeptide of the invention from mammalian tissues can be derived from the sequences provided herein. For a general overview of PCR, *see, Innis et al. PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego* (1990).

[85] Synthetic oligonucleotides can be used to construct genes. This is done using a series of overlapping oligonucleotides, usually 40-120 bp in length, representing

both the sense and anti-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned.

[86] A polynucleotide encoding a polypeptide of the invention can be cloned using intermediate vectors before transformation into mammalian cells for expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors. The proteins can be expressed in either prokaryotes or eukaryotes, using standard methods well known to those of skill in the art.

### III. PURIFICATION OF PROTEINS OF THE INVENTION

[87] Either naturally occurring or recombinant polypeptides of the invention can be purified for use in functional assays. Naturally occurring polypeptides of the invention can be purified from any source (e.g., tissues of an organism expressing an ortholog). Recombinant polypeptides can be purified from any suitable expression system.

[88] The polypeptides of the invention may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (*see, e.g.,* Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Patent No. 4,673,641; Ausubel *et al., supra*; and Sambrook *et al., supra*).

[89] A number of procedures can be employed when recombinant polypeptides are being purified. For example, proteins having established molecular adhesion properties can be reversibly fused to a polypeptide of the invention. With the appropriate ligand, either protein can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein may be then removed by enzymatic activity. Finally polypeptides can be purified using immunoaffinity columns.

#### A. Purification of Proteins from Recombinant Bacteria

[90] When recombinant proteins are expressed by the transformed bacteria in large amounts, typically after promoter induction, although expression can be constitutive, the proteins may form insoluble aggregates. There are several protocols that are suitable for purification of protein inclusion bodies. For example, purification of aggregate proteins (hereinafter referred to as inclusion bodies) typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells typically, but not limited to, by incubation in a buffer of about 100-150 µg/ml lysozyme and 0.1% Nonidet P40, a non-ionic detergent. The cell suspension can be ground using a Polytron grinder (Brinkman



Instruments, Westbury, NY). Alternatively, the cells can be sonicated on ice. Alternate methods of lysing bacteria are described in Ausubel *et al.* and Sambrook *et al.*, both *supra*, and will be apparent to those of skill in the art.

[91] The cell suspension is generally centrifuged and the pellet containing the inclusion bodies resuspended in buffer which does not dissolve but washes the inclusion bodies, *e.g.*, 20 mM Tris-HCl (pH 7.2), 1 mM EDTA, 150 mM NaCl and 2% Triton-X 100, a non-ionic detergent. It may be necessary to repeat the wash step to remove as much cellular debris as possible. The remaining pellet of inclusion bodies may be resuspended in an appropriate buffer (*e.g.*, 20 mM sodium phosphate, pH 6.8, 150 mM NaCl). Other appropriate buffers will be apparent to those of skill in the art.

[92] Following the washing step, the inclusion bodies are solubilized by the addition of a solvent that is both a strong hydrogen acceptor and a strong hydrogen donor (or a combination of solvents each having one of these properties). The proteins that formed the inclusion bodies may then be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to, urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents that are capable of solubilizing aggregate-forming proteins, such as SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of the immunologically and/or biologically active protein of interest. After solubilization, the protein can be separated from other bacterial proteins by standard separation techniques.

[93] Alternatively, it is possible to purify proteins from bacteria periplasm. Where the protein is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to those of skill in the art (*see*, Ausubel *et al.*, *supra*). To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM MgSO<sub>4</sub> and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant

proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

#### **B. Purification of Proteins from Insect Cells**

5                   **[94]** Proteins can also be purified from eukaryotic gene expression systems as described in, e.g., Fernandez and Hoeffler, *Gene Expression Systems* (1999). In some embodiments, baculovirus expression systems are used to isolate proteins of the invention. Recombinant baculoviruses are generally generated by replacing the polyhedrin coding sequence of a baculovirus with a gene to be expressed (e.g., encoding a polypeptide of the invention). Viruses lacking the polyhedrin gene have a unique plaque morphology making them easy to recognize. In some embodiments, a recombinant baculovirus is generated by first cloning a polynucleotide of interest into a transfer vector (e.g., a pUC based vector) such that the polynucleotide is operably linked to a polyhedrin promoter. The transfer vector is transfected with wildtype DNA into an insect cell (e.g., Sf9, Sf21 or BT1-TN-5B1-4 cells),  
10 resulting in homologous recombination and replacement of the polyhedrin gene in the wildtype viral DNA with the polynucleotide of interest. Virus can then be generated and plaque purified. Protein expression results upon viral infection of insect cells. Expressed proteins can be harvested from cell supernatant if secreted, or from cell lysates if intracellular. See, e.g., Ausubel *et al.* and Fernandez and Hoeffler, *supra*.

20

#### **C. Purification of secreted proteins from mammalian cells**

**[95]** Polypeptides of the invention, and in particular, secreted proteins of the invention can be readily purified from mammalian cells expressing the polypeptides. Expression of the polypeptides can be the result of either transient or stable expression of the protein from a recombinant expression cassette introduced into the cells. Secreted proteins  
25 can generally be isolated using standard procedures to purify the proteins from the cell culture medium.

#### **D. Standard Protein Separation Techniques For Purifying Proteins**

30

##### **1. Solubility Fractionation**

**[96]** Often as an initial step, and if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is

ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This will precipitate the most hydrophobic proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, through either dialysis or diafiltration.

Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

## 2. Size Differential Filtration

[97] Based on a calculated molecular weight, a protein of greater and lesser size can be isolated using ultrafiltration through membranes of different pore sizes (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

## 3. Column Chromatography

[98] The proteins of interest can also be separated from other proteins on the basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art.

[99] Immunoaffinity chromatography using antibodies raised to a variety of affinity tags such as hemagglutinin (HA), FLAG, Xpress, Myc, hexahistidine (His), glutathione S transferase (GST) and the like can be used to purify polypeptides. The His tag will also act as a chelating agent for certain metals (e.g., Ni) and thus the metals can also be used to purify His-containing polypeptides. After purification, the tag is optionally removed by specific proteolytic cleavage.

[100] It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

#### IV. DETECTION OF POLYNUCLEOTIDES OF THE INVENTION

[101] Those of skill in the art will recognize that detection of expression of polynucleotides and polypeptides of the invention has many uses. For example, as discussed herein, detection of levels of polynucleotides and polypeptides of the invention in a patient is  
5 useful for diagnosing diabetes or a predisposition for at least some of the pathological effects of diabetes. Moreover, detection of gene expression is useful to identify modulators of expression of polynucleotides and polypeptides of the invention.

[102] A variety of methods of specific DNA and RNA measurement that use nucleic acid hybridization techniques are known to those of skill in the art (*see*, Sambrook,  
10 *supra*). Some methods involve an electrophoretic separation (*e.g.*, Southern blot for detecting DNA, and Northern blot for detecting RNA), but measurement of DNA and RNA can also be carried out in the absence of electrophoretic separation (*e.g.*, by dot blot). Southern blot of genomic DNA (*e.g.*, from a human) can be used for screening for restriction fragment length polymorphism (RFLP) to detect the presence of a genetic disorder affecting a polypeptide of  
15 the invention.

[103] The selection of a nucleic acid hybridization format is not critical. A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins *Nucleic Acid*  
20 *Hybridization, A Practical Approach*, IRL Press (1985); Gall and Pardue, *Proc. Natl. Acad. Sci. U.S.A.*, 63:378-383 (1969); and John *et al. Nature*, 223:582-587 (1969).

[104] Detection of a hybridization complex may require the binding of a signal-generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a  
25 ligand-conjugated probe and an anti-ligand conjugated with a signal. The binding of the signal generation complex is also readily amenable to accelerations by exposure to ultrasonic energy.

[105] The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or antigen, the sample can be detected by  
30 using antibodies. In these systems, a signal is generated by attaching fluorescent or enzyme molecules to the antibodies or in some cases, by attachment to a radioactive label (*see, e.g.*, Tijssen, "*Practice and Theory of Enzyme Immunoassays*," *Laboratory Techniques in*

*Biochemistry and Molecular Biology*, Burdon and van Knippenberg Eds., Elsevier (1985), pp. 9-20).

[106] The probes are typically labeled either directly, as with isotopes, chromophores, lumiphores, chromogens, or indirectly, such as with biotin, to which a streptavidin complex may later bind. Thus, the detectable labels used in the assays of the present invention can be primary labels (where the label comprises an element that is detected directly or that produces a directly detectable element) or secondary labels (where the detected label binds to a primary label, *e.g.*, as is common in immunological labeling). Typically, labeled signal nucleic acids are used to detect hybridization. Complementary nucleic acids or signal nucleic acids may be labeled by any one of several methods typically used to detect the presence of hybridized polynucleotides. The most common method of detection is the use of autoradiography with  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^{32}\text{P}$ -labeled probes or the like.

[107] Other labels include, *e.g.*, ligands that bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies that can serve as specific binding pair members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden *Introduction to Immunocytochemistry*, 2nd ed., Springer Verlag, NY (1997); and in Haugland *Handbook of Fluorescent Probes and Research Chemicals*, a combined handbook and catalogue Published by Molecular Probes, Inc. (1996).

[108] In general, a detector that monitors a particular probe or probe combination is used to detect the detection reagent label. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill in the art. Commonly, an optical image of a substrate comprising bound labeling moieties is digitized for subsequent computer analysis.

[109] The amount of, for example, an RNA is measured by quantifying the amount of label fixed to the solid support by binding of the detection reagent. Typically, the presence of a modulator during incubation will increase or decrease the amount of label fixed to the solid support relative to a control incubation that does not comprise the modulator, or as compared to a baseline established for a particular reaction type. Means of detecting and quantifying labels are well known to those of skill in the art.

[110] In some embodiments, the target nucleic acid or the probe is immobilized on a solid support. Solid supports suitable for use in the assays of the invention

are known to those of skill in the art. As used herein, a solid support is a matrix of material in a substantially fixed arrangement.

[111] A variety of automated solid-phase assay techniques are also appropriate. For instance, very large scale immobilized polymer arrays (VLSIPS™), i.e. Gene Chips or microarrays, available from Affymetrix, Inc. in Santa Clara, CA can be used to detect changes in expression levels of a plurality of genes involved in the same regulatory pathways simultaneously. See, Tijssen, *supra.*, Fodor *et al.* (1991) *Science*, 251: 767- 777; Sheldon *et al.* (1993) *Clinical Chemistry* 39(4): 718-719, and Kozal *et al.* (1996) *Nature Medicine* 2(7): 753-759. Similarly, spotted cDNA arrays (arrays of cDNA sequences bound to nylon, glass or another solid support) can also be used to monitor expression of a plurality of genes.

[112]. Typically, the array elements are organized in an ordered fashion so that each element is present at a specified location on the substrate. Because the array elements are at specified locations on the substrate, the hybridization patterns and intensities (which together create a unique expression profile) can be interpreted in terms of expression levels of particular genes and can be correlated with a particular disease or condition or treatment. See, e.g., Schena *et al.*, *Science* 270: 467-470 (1995)) and (Lockhart *et al.*, *Nature Biotech.* 14: 1675-1680 (1996)).

[113] Hybridization specificity can be evaluated by comparing the hybridization of specificity-control polynucleotide sequences to specificity-control polynucleotide probes that are added to a sample in a known amount. The specificity-control target polynucleotides may have one or more sequence mismatches compared with the corresponding polynucleotide sequences. In this manner, whether only complementary target polynucleotides are hybridizing to the polynucleotide sequences or whether mismatched hybrid duplexes are forming is determined.

[114] Hybridization reactions can be performed in absolute or differential hybridization formats. In the absolute hybridization format, polynucleotide probes from one sample are hybridized to the sequences in a microarray format and signals detected after hybridization complex formation correlate to polynucleotide probe levels in a sample. In the differential hybridization format, the differential expression of a set of genes in two biological samples is analyzed. For differential hybridization, polynucleotide probes from both biological samples are prepared and labeled with different labeling moieties. A mixture of the two labeled polynucleotide probes is added to a microarray. The microarray is then examined under conditions in which the emissions from the two different labels are

individually detectable. Sequences in the microarray that are hybridized to substantially equal numbers of polynucleotide probes derived from both biological samples give a distinct combined fluorescence (Shalon *et al.* PCT publication WO95/35505). In some embodiments, the labels are fluorescent labels with distinguishable emission spectra, such as Cy3 and Cy5 fluorophores.

[115] After hybridization, the microarray is washed to remove nonhybridized nucleic acids and complex formation between the hybridizable array elements and the polynucleotide probes is detected. Methods for detecting complex formation are well known to those skilled in the art. In some embodiments, the polynucleotide probes are labeled with a fluorescent label and measurement of levels and patterns of fluorescence indicative of complex formation is accomplished by fluorescence microscopy, such as confocal fluorescence microscopy.

[116] In a differential hybridization experiment, polynucleotide probes from two or more different biological samples are labeled with two or more different fluorescent labels with different emission wavelengths. Fluorescent signals are detected separately with different photomultipliers set to detect specific wavelengths. The relative abundances/expression levels of the polynucleotide probes in two or more samples are obtained.

[117] Typically, microarray fluorescence intensities can be normalized to take into account variations in hybridization intensities when more than one microarray is used under similar test conditions. In some embodiments, individual polynucleotide probe/target complex hybridization intensities are normalized using the intensities derived from internal normalization controls contained on each microarray.

[118] Detection of nucleic acids can also be accomplished, for example, by using a labeled detection moiety that binds specifically to duplex nucleic acids (*e.g.*, an antibody that is specific for RNA-DNA duplexes). One example uses an antibody that recognizes DNA-RNA heteroduplexes in which the antibody is linked to an enzyme (typically by recombinant or covalent chemical bonding). The antibody is detected when the enzyme reacts with its substrate, producing a detectable product. Coutlee *et al.* (1989) *Analytical Biochemistry* 181:153-162; Bogulavski (1986) *et al. J. Immunol. Methods* 89:123-130; Prooijen-Knegt (1982) *Exp. Cell Res.* 141:397-407; Rudkin (1976) *Nature* 265:472-473, Stollar (1970) *PNAS* 65:993-1000; Ballard (1982) *Mol. Immunol.* 19:793-799; Pisetsky and Caster (1982) *Mol. Immunol.* 19:645-650; Viscidi *et al.* (1988) *J. Clin. Microbiol.* 41:199-209; and Kiney *et al.* (1989) *J. Clin. Microbiol.* 27:6-12 describe antibodies to RNA duplexes,

including homo and heteroduplexes. Kits comprising antibodies specific for DNA:RNA hybrids are available, *e.g.*, from Digene Diagnostics, Inc. (Beltsville, MD).

[119] In addition to available antibodies, one of skill in the art can easily make antibodies specific for nucleic acid duplexes using existing techniques, or modify those antibodies that are commercially or publicly available. In addition to the art referenced above, 5 general methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art (*see, e.g.*, Paul (ed) *Fundamental Immunology, Third Edition* Raven Press, Ltd., NY (1993); Coligan *Current Protocols in Immunology* Wiley/Greene, NY (1991); Harlow and Lane *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY (1989); Stites *et al.* (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, 10 CA, and references cited therein; Goding *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY, (1986); and Kohler and Milstein *Nature* 256: 495-497 (1975)). Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors (*see, Huse et al. Science* 246:1275-1281 (1989); and Ward *et al. Nature* 341:544-546 (1989)). Specific monoclonal and 15 polyclonal antibodies and antisera will usually bind with a  $K_D$  of at least about 0.1  $\mu M$ , preferably at least about 0.01  $\mu M$  or better, and most typically and preferably, 0.001  $\mu M$  or better.

[120] The nucleic acids used in this invention can be either positive or 20 negative probes. Positive probes bind to their targets and the presence of duplex formation is evidence of the presence of the target. Negative probes fail to bind to the suspect target and the absence of duplex formation is evidence of the presence of the target. For example, the use of a wild type specific nucleic acid probe or PCR primers may serve as a negative probe in an assay sample where only the nucleotide sequence of interest is present.

[121] The sensitivity of the hybridization assays may be enhanced through 25 use of a nucleic acid amplification system that multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario) and Q 30 Beta Replicase systems. These systems can be used to directly identify mutants where the PCR or LCR primers are designed to be extended or ligated only when a selected sequence is present. Alternatively, the selected sequences can be generally amplified using, for example, nonspecific PCR primers and the amplified target region later probed for a specific sequence indicative of a mutation. It is understood that various detection probes, including Taqman



and molecular beacon probes can be used to monitor amplification reaction products, e.g., in real time.

[122] An alternative means for determining the level of expression of the nucleic acids of the present invention is *in situ* hybridization. *In situ* hybridization assays are well known and are generally described in Angerer *et al.*, *Methods Enzymol.* 152:649-660 (1987). In an *in situ* hybridization assay, cells, preferentially human cells from the cerebellum or the hippocampus, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

[123] Single nucleotide polymorphism (SNP) analysis is also useful for detecting differences between alleles of the polynucleotides (e.g., genes) of the invention. SNPs linked to genes encoding polypeptides of the invention are useful, for instance, for diagnosis of diseases (e.g., diabetes) whose occurrence is linked to the gene sequences of the invention. For example, if an individual carries at least one SNP linked to a disease-associated allele of the gene sequences of the invention, the individual is likely predisposed for one or more of those diseases. If the individual is homozygous for a disease-linked SNP, the individual is particularly predisposed for occurrence of that disease (e.g., diabetes). In some embodiments, the SNP associated with the gene sequences of the invention is located within 300,000; 200,000; 100,000; 75,000; 50,000; or 10,000 base pairs from the gene sequence.

[124] Various real-time PCR methods including, e.g., Taqman or molecular beacon-based assays (e.g., U.S. Patent Nos. 5,210,015; 5,487,972; Tyagi *et al.*, *Nature Biotechnology* 14:303 (1996); and PCT WO 95/13399 are useful to monitor for the presence of absence of a SNP. Additional SNP detection methods include, e.g., DNA sequencing, sequencing by hybridization, dot blotting, oligonucleotide array (DNA Chip) hybridization analysis, or are described in, e.g., U.S. Patent No. 6,177,249; Landegren *et al.*, *Genome Research*, 8:769-776 (1998); Botstein *et al.*, *Am J Human Genetics* 32:314-331 (1980); Meyers *et al.*, *Methods in Enzymology* 155:501-527 (1987); Keen *et al.*, *Trends in Genetics* 7:5 (1991); Myers *et al.*, *Science* 230:1242-1246 (1985); and Kwok *et al.*, *Genomics* 23:138-144 (1994).

## V. DETECTION OF POLYPEPTIDES OF THE INVENTION

[125] In addition to the detection of polynucleotides of the invention and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect polypeptides of the invention. Immunoassays can be used to qualitatively or quantitatively analyze polypeptides of the invention. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

### A. Antibodies to Target Proteins or other immunogens

[126] Methods for producing polyclonal and monoclonal antibodies that react specifically with a protein of interest or other immunogen are known to those of skill in the art (*see, e.g.*, Coligan, *supra*; and Harlow and Lane, *supra*; Stites *et al.*, *supra* and references cited therein; Goding, *supra*; and Kohler and Milstein *Nature*, 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors (*see, Huse et al.*, *supra*; and Ward *et al.*, *supra*). For example, in order to produce antisera for use in an immunoassay, the protein of interest or an antigenic fragment thereof, is isolated as described herein. For example, a recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[127] Polyclonal sera are collected and titered against the immunogen in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of  $10^4$  or greater are selected and tested for their crossreactivity against proteins other than the polypeptides of the invention or even other homologous proteins from other organisms, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a  $K_D$  of at least about 0.1 mM, more usually at least about 1  $\mu$ M, preferably at least about 0.1  $\mu$ M or better, and most preferably, 0.01  $\mu$ M or better.

[128] A number of proteins of the invention comprising immunogens may be used to produce antibodies specifically or selectively reactive with the proteins of interest. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure form. Synthetic peptides made using the protein sequences described herein may also be

used as an immunogen for the production of antibodies to the protein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells and purified as generally described *supra*. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

[129] Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to polypeptides of the invention. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see*, Harlow and Lane, *supra*).

[130] Monoclonal antibodies may be obtained using various techniques familiar to those of skill in the art. Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*see*, Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include, *e.g.*, transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences that encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *supra*.

[131] Once target immunogen-specific antibodies are available, the immunogen can be measured by a variety of immunoassay methods with qualitative and quantitative results available to the clinician. For a review of immunological and immunoassay procedures in general *see*, Stites, *supra*. Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Maggio *Enzyme Immunoassay*, CRC Press, Boca Raton, Florida (1980); Tijssen, *supra*; and Harlow and Lane, *supra*.

[132] Immunoassays to measure target proteins in a human sample may use a polyclonal antiserum that was raised to full-length polypeptides of the invention or a fragment thereof. This antiserum is selected to have low cross-reactivity against other proteins and any such cross-reactivity is removed by immunoabsorption prior to use in the immunoassay.

#### **B. Immunological Binding Assays**

[133] In some embodiments, a protein of interest is detected and/or quantified using any of a number of well-known immunological binding assays (*see, e.g., U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168*). For a review of the general immunoassays, *see also Asai Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Academic Press, Inc. NY (1993); Stites, *supra*. Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (*e.g., full-length polypeptides of the present invention, or antigenic subsequences thereof*). The capture agent is a moiety that specifically binds to the analyte. The antibody may be produced by any of a number of means well known to those of skill in the art and as described above.

[134] Immunoassays also often utilize a labeling agent to bind specifically to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/protein complex.

[135] In a preferred embodiment, the labeling agent is a second antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second antibody can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

[136] Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, can also be used as the label agents. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of

species (see, generally, Kronval, *et al. J. Immunol.*, 111:1401-1406 (1973); and Akerstrom, *et al. J. Immunol.*, 135:2589-2542 (1985)).

[137] Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. The incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

#### 1. Non-Competitive Assay Formats

[138] Immunoassays for detecting proteins or analytes of interest from tissue samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured protein or analyte is directly measured. In one preferred "sandwich" assay, for example, the capture agent (e.g., antibodies specific for the polypeptides of the invention) can be bound directly to a solid substrate where it is immobilized. These immobilized antibodies then capture the polypeptide present in the test sample. The polypeptide of the invention thus immobilized is then bound by a labeling agent, such as a second labeled antibody specific for the polypeptide. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

#### 2. Competitive Assay Formats

[139] In competitive assays, the amount of protein or analyte present in the sample is measured indirectly by measuring the amount of an added (exogenous) protein or analyte displaced (or competed away) from a specific capture agent (e.g., antibodies specific for a polypeptide of the invention) by the protein or analyte present in the sample. The amount of immunogen bound to the antibody is inversely proportional to the concentration of immunogen present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of analyte may be detected by providing a labeled analyte molecule. It is understood that labels can include, e.g., radioactive labels as well as peptide or other tags that can be recognized by detection reagents such as antibodies.

[140] Immunoassays in the competitive binding format can be used for cross-reactivity determinations. For example, the protein encoded by the sequences described herein can be immobilized on a solid support. Proteins are added to the assay and compete

with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to that of the protein encoded by any of the sequences described herein. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the considered proteins, *e.g.*, distantly related homologs.

[141] The immunoabsorbed and-pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps a protein of the present invention, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than 10 times the amount of the protein partially encoded by a sequence herein that is required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the target protein.

### 3. Other Assay Formats

[142] In some embodiments, western blot (immunoblot) analysis is used to detect and quantify the presence of a polypeptide of the invention in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support (such as, *e.g.*, a nitrocellulose filter, a nylon filter, or a derivatized nylon filter) and incubating the sample with the antibodies that specifically bind the protein of interest. For example, antibodies are selected that specifically bind to the polypeptides of the invention on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the antibodies against the protein of interest.

[143] Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (*see, Monroe et al. (1986) Amer. Clin. Prod. Rev. 5:34-41*).

### 4. Labels

[144] The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific

binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most labels useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (*e.g.*, Dynabeads<sup>TM</sup>), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*, <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, *etc.*) beads.

[145] The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on the sensitivity required, the ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

[146] Non-radioactive labels are often attached by indirect means. The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorescent compound. A variety of enzymes and fluorescent compounds can be used with the methods of the present invention and are well-known to those of skill in the art (for a review of various labeling or signal producing systems which may be used, *see, e.g.*, U.S. Patent No. 4,391,904).

[147] Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected directly by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

[148] Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In

this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need to be labeled and the presence of the target antibody is detected by simple visual inspection.

## **5 VI. IDENTIFICATION OF MODULATORS OF POLYPEPTIDES OF THE INVENTION**

[149] Modulators of a polypeptide of the invention, i.e. agonists or antagonists of a polypeptide's activity, or polypeptide's or polynucleotide's expression or full-length polypeptides of the invention or fragments thereof, are useful for treating a number of human diseases, including diabetes or obesity. For example, administration of modulators can be used to treat diabetic patients or prediabetic individuals to prevent progression, and therefore symptoms, associated with diabetes (including insulin resistance). Modulators of the invention can also be used to reduce obesity as well as the various diseases associated with obesity (e.g., gallbladder disease, cancer, sleep apnea, atherosclerosis, diabetes, and hypertension). In some cases, the modulators of the invention are used to regulate body physiology to reduce the chance of obesity-related diseases. For example, the modulators can be used to regulate serum lipids (total cholesterol, low-density lipoprotein (LDL), cholesterol, LDL/high density lipoprotein ratio and triglycerides).

### **A. Agents that Modulate Polypeptides of the Invention**

[150] The agents tested as modulators of polypeptides of the invention can be any small chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. Modulators include agents designed to reduce the level of mRNA encoding a polypeptide of the invention (e.g. antisense molecules, ribozymes, DNazymes, small inhibitory RNAs and the like) or the level of translation from an mRNA (e.g., translation blockers such as an antisense molecules that are complementary to translation start or other sequences on an mRNA molecule). Modulators of the invention also include antibodies that specific bind to and/or inhibit or activate the polypeptides of the invention. Other modulators include the polypeptides of the invention themselves, fragments thereof, or fusion proteins comprising the polypeptides or fragments thereof (e.g., in some embodiments, comprising at least 25, 50, or 100 amino acids of the polypeptide). For polypeptides of the invention that are receptors, soluble fragments of the polypeptides (i.e., lacking a transmembrane domain) can act as modulators of polypeptide



signaling activity. For polypeptides of the invention that are secreted, both full length and fragments with biological activity can act as modulators. It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

[151] In some embodiments, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

[152] A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[153] Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.*, U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (*e.g.*, PCT Publication No. WO 91/19735), encoded peptides (*e.g.*, PCT Publication WO 93/20242), random bio-oligomers (*e.g.*, PCT Publication No. WO 92/00091), benzodiazepines (*e.g.*, U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see* Ausubel, Berger and Sambrook, *all supra*), peptide nucleic acid

libraries (*see, e.g.*, U.S. Patent 5,539,083), antibody libraries (*see, e.g.*, Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.*, Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (*see, e.g.*, benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, 5,288,514, and the like).

[154] Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.*, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.*, ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, MO, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, etc.).

#### **B. Methods of Screening for Modulators of the Polypeptides of the Invention**

[155] A number of different screening protocols can be utilized to identify agents that modulate the level of expression or activity of a polynucleotide of a polypeptide of the invention in cells, particularly mammalian cells, and especially human cells. In general terms, the screening methods involve screening a plurality of agents to identify an agent that modulates the activity of a polypeptide of the invention by, *e.g.*, binding to the polypeptide, preventing an inhibitor or activator from binding to the polypeptide, increasing association of an inhibitor or activator with the polypeptide, or activating or inhibiting expression of the polypeptide. The assays can be designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays).

[156] Any cell expressing a full-length polypeptide of the invention or a fragment thereof can be used to identify modulators. In some embodiments, the cells are eukaryotic cells lines (*e.g.*, CHO or HEK293) transformed to express a heterologous polypeptide of the invention. In some embodiments, a cell expressing an endogenous polypeptide of the invention is used in screens. In other embodiments, modulators are screened for their ability to affect insulin responses. In other embodiments, modulators are screened for their ability to effect body weight (as measured by BMI or waist-to-hip ratio)

and secretion of a variety of obesity markers (e.g., leptin, IL-6 or TNF alpha). In other embodiments, modulators are screened for their ability to effect lipid metabolism. In other embodiments, modulators are screened for their ability to effect the secretion and activity of adipogenic factors.

5           [157] In some embodiments, modulators of ADL1CAN comprising the amino acid sequence of SEQ ID NO: 2, 4, or 6, may be identified using, e.g., modulator binding assays, expression assays or promoter-reporter assays.

          [158] In some embodiments, modulators of ALDH1A3 comprising the amino acid sequence of SEQ ID NO: 8, 10, or 12, may be identified using, e.g., modulator  
10 binding assays, expression assays, promoter-reporter assays, or assays based on the retinoic acid production.

          [159] In some embodiments, modulators of ALK7 comprising the amino acid sequence of SEQ ID NO: 14, 16, or 18, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays or kinase assays based on Smad2  
15 phosphorylation. Kinase assays can be carried out after contacting either purified recombinant ALK7 protein, or an intact cell with the modulator. Modulators which bind to the ALK7 can be screened by a ligand binding assay method using e.g. nodal as the ligand.

          [160] In some embodiments, modulators of C3AR1 comprising the amino acid sequence of SEQ ID NO: 20, 22, or 24, may be identified using, e.g., the expression  
20 assays, promoter-reporter assays, binding assays, or screening methods that monitor modulator-induced fluctuation of intracellular  $Ca^{++}$  concentration. Modulators which bind to the C3AR1 can be screened by a ligand binding assay method using e.g. complement anaphylatoxin C3a as the ligand.

          [161] In some embodiments, modulators of CALCRL comprising the amino acid sequence of SEQ ID NO: 26, 28, or 30, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor  
25 modulator-induced fluctuation of intracellular cAMP concentration. Modulators which bind to the CALCRL can be screened by a ligand binding assay method using, e.g., adrenomedullin or calcitonin gene related peptide as ligands.

30           [162] In some embodiments, modulators of CCL13 comprising the amino acid sequence of SEQ ID NO: 32, 33, 35, or 37, may be identified using, e.g., expression assays promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation of intracellular  $Ca^{++}$  concentration. Modulators

which bind to the CCL13 can be screened by a ligand binding assay method using, e.g., CCR1 or other C-C G-protein coupled receptors known to bind to CCL13.

5 [163] In some embodiments, modulators of CCL8 comprising the amino acid sequence of SEQ ID NO: 39, 40, 42, or 44, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation of intracellular  $\text{Ca}^{++}$  concentration. Modulators which bind to the CCL8 can be screened by a ligand binding assay method using, e.g., CCR1 or other C-C G-protein coupled receptors known to bind to CCL8.

10 [164] In some embodiments, modulators of CHI3L1 comprising the amino acid sequence of SEQ ID NO: 46, 47, 49, or 51, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation of phosphorylation and activity of MAPK and/or AKT (*see, e.g., Recklies, A.D. et al., Biochem J. 365:119-26 (2002)*).

15 [165] In some embodiments, modulators of CR1 comprising the amino acid sequence of SEQ ID NO: 53 or 55, may be identified using, e.g., expression assays, promoter-reporter assays, or modulator binding assays. Modulators which bind to the CR1 can be screened by a ligand binding assay method using, e.g., complement component C3b as ligand.

20 [166] In some embodiments, modulators of CSFR1 comprising the amino acid sequence of SEQ ID NO: 57, 59 or 61, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays or activity assays. Kinase assays can be carried out after contacting either purified recombinant CSFR1 protein or an intact cell with modulators. Modulators which bind to the CSFR1 can be screened by a ligand binding assay method using e.g clony stimulating factor as ligand.

25 [167] In some embodiments, modulators of CTSK comprising the amino acid sequence of SEQ ID NO: 63, 64, 66 or 68, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or enzyme assays. Enzyme assays can be carried out after contacting either purified recombinant CTSK protein, or an intact cell with a modulator using e.g. fibrinogen as a substrate.

30 [168] In some embodiments, modulators of CXCR4 comprising the amino acid sequence of SEQ ID NO: 70, 72 or 74, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation of intracellular  $\text{Ca}^{++}$  concentration or phosphorylation and

activity of MAPK and/or AKT. Modulators which bind to the CXCR4 can be screened by a ligand binding assay method using e.g. CXCL12 as a ligand.

5 [169] In some embodiments, modulators of DDAH2 comprising the amino acid sequence of SEQ ID NO: 76, 78 or 80, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or activity assays. Modulators which effect DDAH2 activity can be screened by measuring the conversation of ADMA to citrulline and methylamines.

10 [170] In some embodiments, modulators of DERP7 comprising the amino acid sequence of SEQ ID NO: 82, 84 or 86, may be identified using, e.g., expression assays, modulator binding assays, or promoter-reporter assays.

[171] In some embodiments, modulators of ENDOGLYX1 comprising the amino acid sequence of SEQ ID NO: 88, 90 or 92, may be identified using, e.g., expression assays, modulator binding assays, promoter-reporter or activity assays based on angiogenesis (*see, e.g.,* Christian, S. *et al., J. Biol. Chem.* 276: 48588-48595 (2001)).

15 [172] In some embodiments, modulators of ETL comprising the amino acid sequence of SEQ ID NO: 94, 96 or 98, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or assays based on a G-protein coupled receptor activity.

20 [173] In some embodiments, modulators of FLJ12389 comprising the amino acid sequence of SEQ ID NO: 100, 102, 104, 106 or 108, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or assays based on an AMP binding or an acetoacetate-CoA ligase activity.

25 [174] In some embodiments, modulators of FZD4 comprising the amino acid sequence of SEQ ID NO: 110, 112, 114 or 116, may be identified using, e.g., expression assays, promoter-reporter assays or modulator binding assays. Modulators which bind to the FZD4 can be screened by a ligand binding assay method using e.g. norrin as a ligand.

30 [175] In some embodiments, modulators of GLIPR1 comprising the amino acid sequence of SEQ ID NO: 118, 120 or 122, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or activity assays based on an induction of apoptosis.

[176] In some embodiments, modulators of GPR105 comprising the amino acid sequence of SEQ ID NO: 124, 126 or 128, may be identified using, e.g., expression assays, promoter-reporetr assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation of intracellular Ca<sup>++</sup> concentration. Modulators

which bind to the GPR105 can be screened by a ligand binding assay method using, e.g., UDP-glucose as ligand.

[177] In some embodiments, modulators of GPR146 comprising the amino acid sequence of SEQ ID NO: 130, 132 or 134, may be identified using, e.g., expression  
5 assays, promoter-reporter assays, modulator binding assays, or assays based on a G-protein coupled receptor activity.

[178] In some embodiments, modulators of GPR30 comprising the amino acid sequence of SEQ ID NO: 136, 138 or 140, may be identified using, e.g., expression  
10 assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation of intracellular cAMP concentration or phosphorylation and activity of MAPK.

[179] In some embodiments, modulators of GPR65 comprising the amino acid sequence of SEQ ID NO: 142, 144 or 146, may be identified using, e.g., expression  
15 assays, promoter-reporter assays or modulator binding assays. Modulators which bind to the GPR65 can be screened by a ligand binding assay method using, e.g., psychosine as ligand.

[180] In some embodiments, modulators of HTR2B comprising the amino acid sequence of SEQ ID NO: 148, 150 or 152, may be identified using, e.g., expression  
20 assays, promoter-reporter assays or modulator binding assays. Modulators which bind to the HTR2B can be screened by a ligand binding assay method using, e.g., serotonin as ligand. Assays detecting phosphoinositide phospholipase C activity can be used.

[181] In some embodiments, modulators of ITGB2 comprising the amino acid sequence of SEQ ID NO: 154, 156 or 158, may be identified using, e.g., expression  
25 assays, promoter-reporter assays or modulator binding assays. Modulators which bind to the ITGB2 can be screened by a ligand binding assay method using, e.g., ITG alpha chain protein.

[182] In some embodiments, modulators of ITIH5 comprising the amino acid sequence of SEQ ID NO: 160, 161 or 163, may be identified using, e.g., expression assays,  
promoter-reporter assays, or modulator binding assays.

[183] In some embodiments, modulators of LGALS12 comprising the amino acid sequence of SEQ ID NO: 165, 167, 169, 171, 173, 175, 177 or 179, may be identified  
30 using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation in apoptosis.

[184] In some embodiments, modulators of NMB comprising the amino acid sequence of SEQ ID NO: 181, 182, 184 or 186, may be identified using, e.g., expression  
assays, promoter-reporter assays, modulator binding assays, screening methods that monitor

modulator-induced fluctuation of intracellular  $\text{Ca}^{++}$  concentration or binding assays.

Modulators which bind to the NMB can be screened by a ligand binding assay method using e.g. NMBR.

5 [185] In some embodiments, modulators of NNAT comprising the amino acid sequence of SEQ ID NO: 188, 190, 192 or 194, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays.

[186] In some embodiments, modulators of OLFM2 comprising the amino acid sequence of SEQ ID NO: 196, 197, 199 or 201, may be identified using, e.g., expression assays, promoter-reporter assays, or modulator binding assays.

10 [187] In some embodiments, modulators of OPN3 comprising the amino acid sequence of SEQ ID NO: 203, 205, 207, 209 or 211, may be identified using, e.g., expression assays, promoter-reporter assays, or modulator binding assays.

[188] In some embodiments, modulators of PTPRE comprising the amino acid sequence of SEQ ID NO: 213, 215, 217, 219 or 221, may be identified using, e.g.,  
15 expression assays, promoter-reporter assays, modulator binding assays, or screening assays based on a receptor protein tyrosine phosphatase activity or phosphorylation and activity of MAPK.

[189] In some embodiments, modulators of RDC1 comprising the amino acid sequence of SEQ ID NO: 223, 225 or 227, may be identified using, e.g., expression assays,  
20 promoter-reporter assays, or modulator binding assays.

[190] In some embodiments, modulators of SLIT2 comprising the amino acid sequence of SEQ ID NO: 229, 230, 232 or 234, may be identified using, e.g., expression assays, promoter-reporter assays or modulator binding assays. Modulators which bind to the SLIT2 can be screened by a ligand binding assay method using, e.g., roundabout receptor  
25 ROBO1.

[191] In some embodiments, modulators of TNFRSF21 comprising the amino acid sequence of SEQ ID NO: 236, 238 or 240, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation in apoptosis and activation of both NF-kappaB  
30 and JNK.

[192] In some embodiments, modulators of TNFSF13B comprising the amino acid sequence of SEQ ID NO: 242, 243, 245, 247 or 249, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays based on a receptor

activity using e.g. TNFRSF13b, TNFRSF13c or TNFRSF17 or screening methods that monitor modulator-induced fluctuation in activation of NF-kappaB.

[193] In some embodiments, modulators of TNFSF14 comprising the amino acid sequence of SEQ ID NO: 251, 252, 254, 256, 258 or 260, may be identified using, e.g., expression assays, promoter-reporter assays, or modulator binding assays based on a receptor activity using e.g. TNFRSF14.

[194] In some embodiments, modulators of TPSB2 comprising the amino acid sequence of SEQ ID NO: 262, 263, 265 or 267, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation in serine-type peptidase activity.

[195] In some embodiments, modulators of WISP2 comprising the amino acid sequence of SEQ ID NO: 269, 270, 272 or 274, may be identified using, e.g., expression assays, promoter-reporter assays, modulator binding assays, or screening methods that monitor modulator-induced fluctuation in proliferative rate of vascular smooth muscle cells.

15

#### 1. Polypeptide Binding Assays

[196] Preliminary screens can be conducted by screening for agents capable of binding to polypeptides of the invention, as at least some of the agents so identified are likely modulators of a polypeptide of the invention. Binding assays are also useful, e.g., for identifying endogenous proteins that interact with polypeptides of the invention. For example, antibodies, receptors or other molecules that bind polypeptides of the invention can be identified in binding assays.

[197] Binding assays usually involve contacting a polypeptide of the invention with one or more test agents and allowing sufficient time for the protein and test agents to form a binding complex. Any binding complexes formed can be detected using any of a number of established analytical techniques. Protein binding assays include, but are not limited to, methods that measure co-precipitation or co-migration on non-denaturing SDS-polyacrylamide gels, and co-migration on Western blots (*see, e.g.*, Bennet, J.P. and Yamamura, H.I. (1985) "Neurotransmitter, Hormone or Drug Receptor Binding Methods," in *Neurotransmitter Receptor Binding* (Yamamura, H. I., *et al.*, eds.), pp. 61-89. Other binding assays involve the use of mass spectrometry or NMR techniques to identify molecules bound to a polypeptide of the invention or displacement of labeled substrates. The polypeptides of the invention utilized in such assays can be naturally expressed, cloned or synthesized.



[198] In addition, mammalian or yeast two-hybrid approaches (*see, e.g.,* Bartel, P.L. *et. al. Methods Enzymol*, 254:241 (1995)) can be used to identify polypeptides or other molecules that interact or bind when expressed together in a host cell.

5                   2.     Polypeptide Activity

[199] The activity of polypeptides of the invention can be assessed using a variety of *in vitro* and *in vivo* assays to determine functional, chemical, and physical effects, *e.g.,* measuring ligand binding (*e.g.,* radioactive or otherwise labeled ligand binding), second messengers (*e.g.,* cAMP, cGMP, IP<sub>3</sub>, DAG, or Ca<sup>2+</sup>), ion flux, phosphorylation levels, transcription levels, and the like. Measurement of such functional, chemical and/or physical effects may be direct (*e.g.,* directly detecting calcium flux) or indirect (*e.g.,* detecting changes in expression or activity of gene products that are known to be modulated by the effects such as calcium flux or others listed above). Furthermore, such assays can be used to test for inhibitors and activators of the polypeptides of the invention. Modulators can also be  
10                   genetically altered versions of polypeptides of the invention.  
15

[200] The polypeptide of the assay will be selected from a polypeptide with substantial identity to a sequence of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272, 274 or other conservatively modified variants thereof.  
20                   Generally, the amino acid sequence identity will be at least 70%, optionally at least 85%, optionally at least 90, or optionally at least 95% to the polypeptides exemplified herein. Optionally, the polypeptide of the assays will comprise a fragment of a polypeptide of the invention, such as an extracellular domain, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. Either a  
25                   polypeptide of the invention or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.  
30

[201] Modulators of polypeptide activity are tested using either recombinant or naturally occurring polypeptides of the invention. The protein can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal,

either recombinant or naturally occurring. For example, tissue slices, dissociated cells, e.g., from tissues expressing polypeptides of the invention, transformed cells, or membranes can be used. Modulation is tested using one of the *in vitro* or *in vivo* assays described herein.

[202] Modulator binding to polypeptides of the invention, a domain, or chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape), chromatographic, or solubility properties.

[203] Samples or assays that are treated with a potential modulator (e.g., a "test compound") are compared to control samples without the test compound, to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative activity value of 100. Inhibition of the polypeptides of the invention is achieved when the activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of the polypeptides of the invention is achieved when the activity value relative to the control is 110%, optionally 150%, 200%, 300%, 400%, 500%, or 1000-2000%.

### 3. Expression Assays

[204] Screening for a compound that modulates the expression of a polynucleotide or a polypeptide of the invention is also provided. Screening methods generally involve conducting cell-based assays in which test compounds are contacted with one or more cells expressing a polynucleotide or a polypeptide of the invention, and then detecting an increase or decrease in expression (either transcript or translation product). Assays can be performed with any cells that express a polynucleotide or a polypeptide of the invention.

[205] Expression can be detected in a number of different ways. As described *infra*, the expression level of a polynucleotide of the invention in a cell can be determined by probing the mRNA expressed in a cell with a probe that specifically hybridizes with a transcript (or complementary nucleic acid derived there from) of a polynucleotide of the invention. Probing can be conducted by lysing the cells and conducting Northern blots or without lysing the cells using *in situ*-hybridization techniques. Alternatively, a polypeptide of the invention can be detected using immunological methods in which a cell lysate is probed with antibodies that specifically bind to the polypeptide.

[206] Promoter-reporter assays can be carried out using mammalian cells transfected with a reporter gene operably linked to sequences derived from the promoter

regions of genes encoding the polypeptides of the invention. The increased or decreased expression of the reporter gene can be detected in the presence and absence of the modulator. Expression of reporter genes may be detected by hybridization to a complementary nucleic acid, by using an immunological reagent, by assaying for an activity of the reporter gene product, or other methods known to those in the art .

[207] The level of expression or activity of a polynucleotide or a polypeptide of the invention can be compared to a baseline value. The baseline value can be a value for a control sample or a statistical value that is representative of expression levels of a polynucleotide or a polypeptide of the invention for a control population (e.g., lean individuals as described herein) or cells (e.g., tissue culture cells not exposed to a modulator). Expression levels can also be determined for cells that do not express the polynucleotide or a polypeptide of the invention as a negative control. Such cells generally are otherwise substantially genetically the same as the test cells.

[208] A variety of different types of cells can be utilized in the reporter assays. Cells that do not endogenously express a polypeptide of the invention can be prokaryotic, but are preferably eukaryotic. The eukaryotic cells can be any of the cells typically utilized in generating cells that harbor recombinant nucleic acid constructs. Exemplary eukaryotic cells include, but are not limited to, yeast, and various higher eukaryotic cells such as the HEK293, HepG2, COS, CHO and HeLa cell lines.

[209] Various controls can be conducted to ensure that an observed activity is authentic including running parallel reactions with cells that lack the reporter construct or by not contacting a cell harboring the reporter construct with test compound. Compounds can also be further validated as described below.

#### 4. Validation

[210] Agents that are initially identified by any of the foregoing screening methods can be further tested to validate the apparent activity. Alternatively, potential modulators can be tested initially using the foregoing validation assays without preliminary screening.

[211] Modulators that are selected for further study can be tested for anti-diabetic effects using the "classic" insulin responsive cell line, mouse 3T3-L1 adipocytes, muscle cells such as L6 cells and the like. Cells (e.g., adipocytes or muscle cells) are pre-incubated with the modulators and tested for acute (up to 4 hours) and chronic (overnight) effects on basal and insulin-stimulated GLUT4 translocation and glucose uptake.

[212] Modulators that are selected for further study can be tested for anti-obesity effects using any adipocyte or adipogenic cell, e.g., mouse cell line 3T3-L1 adipocytes, freshly isolated rodent or human adipocytes, undifferentiated adipogenic cells and the like. Cells (e.g., adipocytes cells) are pre-incubated with the modulators and tested for acute (up to 4 hours) and chronic (overnight or longer) effects on basal and insulin-stimulated release of adipogenic factors, adipocyte cell size, leptin and TNF alpha release, and/or lipid metabolism. Undifferentiated adipogenic cells can be pre-incubated with the modulators and tested for effects on differentiation into adipocytes (including changes in differentiation markers) and/or triglyceride accumulation.

[213] The response of this increase in body weight can be determined at an organismal, tissue or cellular level. For example, increased fasting blood leptin levels are indicative of obesity. Other methods of measuring obesity include, e.g., calculation of BMI, waist-to-hip ratio, total body fat, measuring the blood levels of a variety of secreted proteins which have been shown to correlate to obesity (IL-6, TNF alpha) and measuring the fasted blood levels of free fatty acids.

[214] Following such studies, validity of the modulators is tested in suitable animal models. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model for humans and then determining if expression of activity of a polypeptide of the invention is in fact modulated.

[215] The effect of the compound will be assessed in either obese animals, diabetic animals or in diet induced insulin resistant animals. The body weight loss, blood glucose and insulin levels will be determined. The animal models utilized in validation studies generally are mammals of any kind. Specific examples of suitable animals include, but are not limited to, primates, mice and rats. Monogenic models of diabetes (e.g., ob/ob and db/db mice, Zucker rats and Zucker Diabetic Fatty rats, etc.) or polygenic models of diabetes (e.g., OLETF rats, GK rats, NSY mice, and KK mice) can be useful for validating modulation of a polypeptide of the invention in a diabetic or insulin resistant animal. In addition, transgenic animals expressing human polypeptides of the invention can be used to further validate drug candidates.

[216] Monogenic models of obesity (e.g., OLETF, tubby, mahogany, agouti, ob/ob and db/db mice etc) or polygenic models of obesity (e.g., high fat diet-induced obese animals, NZO mice, KK mice, Wellesley mice, GK rats, etc.) can be useful for validating modulation of a polypeptide of the invention in an obese animal. The most widely used criteria for assessing the efficacy of anti-obesity treatments are those from the FDA. The

FDA defines a body weight loss of >5% as statistically significant compared to placebo. However, it will be appreciated that any detectable change in body weight following administration of a modulator of the invention can be considered a relevant result.

5           **C.     Solid Phase and Soluble High Throughput Assays**

          [217] In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells  
10 can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (*e.g.*, 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 or more different compounds are possible using the integrated systems of the invention. In addition, microfluidic approaches to reagent  
15 manipulation can be used.

          [218] A molecule of interest (*e.g.*, a polypeptide or polynucleotide of the invention, or a modulator thereof) can be bound to the solid-state component, directly or indirectly, via covalent or non-covalent linkage, *e.g.*, via a tag. The tag can be any of a variety of components. In general, a molecule that binds the tag (a tag binder) is fixed to a  
20 solid support, and the tagged molecule of interest is attached to the solid support by interaction of the tag and the tag binder.

          [219] A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with  
25 appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, poly-His, *etc.*) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (*see*, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

          [220] Similarly, any haptenic or antigenic compound can be used in  
30 combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody that recognizes the first antibody. In addition to antibody-

antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs, such as agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherin  
5 family, the integrin family, the selectin family, and the like; see, e.g., Pigott & Power, *The Adhesion Molecule Facts Book I* (1993)). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g., which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations),  
10 oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

[221] Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder  
15 pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

[222] Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly-gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to those of skill in the art. For  
20 example, poly(ethylene glycol) linkers are available from Shearwater Polymers, Inc., Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

[223] Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by  
25 exposing all or a portion of the substrate to a chemical reagent that fixes a chemical group to the surface that is reactive with a portion of the tag binder. For example, groups that are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer  
30 arrays is well described in the literature (see, e.g., Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, e.g., peptides); Geysen *et al.*, *J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank and Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al.*, *Science*, 251:767-777 (1991); Sheldon *et al.*,

*Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al.*, *Nature Medicine* 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

5           [224] The invention provides *in vitro* assays for identifying, in a high throughput format, compounds that can modulate the expression or activity of a polypeptide of the invention. Control reactions that measure activity of a polypeptide of the invention in a cell in a reaction that does not include a potential modulator are optional, as the assays are highly uniform. Such optional control reactions are appropriate and increase the reliability of  
10 the assay. Accordingly, in some embodiments, the methods of the invention include such a control reaction. For each of the assay formats described, "no modulator" control reactions that do not include a modulator provide a background level of binding activity.

          [225] In some assays it will be desirable to have positive controls. At least two types of positive controls are appropriate. First, a known activator of a polypeptide or a  
15 polynucleotide of the invention can be incubated with one sample of the assay, and the resulting increase in signal resulting from an increased expression level or activity of a polypeptide or a polynucleotide of the invention are determined according to the methods herein. Second, a known inhibitor of a polypeptide or a polynucleotide of the invention can be added, and the resulting decrease in signal for the expression or activity of a polypeptide  
20 or a polynucleotide of the invention can be similarly detected. It will be appreciated that modulators can also be combined with activators or inhibitors to find modulators that inhibit the increase or decrease that is otherwise caused by the presence of the known modulator of a polypeptide or a polynucleotide of the invention.

## VII. COMPOSITIONS, KITS AND INTEGRATED SYSTEMS

25           [226] The invention provides compositions, kits and integrated systems for practicing the assays described herein using nucleic acids or polypeptides of the invention, antibodies, etc.

          [227] The invention provides assay compositions for use in solid phase assays; such compositions can include, for example, one or more nucleic acids encoding a  
30 polypeptide of the invention immobilized on a solid support, and a labeling reagent. In each case, the assay compositions can also include additional reagents that are desirable for hybridization. Modulators of expression or activity of a polypeptide of the invention can also be included in the assay compositions.

[228] The invention also provides kits for carrying out the assays of the invention. The kits typically include a probe that comprises (1) an antibody that specifically binds to a polypeptide of the invention or (2) a polynucleotide sequence encoding at least a fragment of such polypeptides, and a label for detecting the presence of the probe. The kits  
5 may include at least one polynucleotide sequence encoding a polypeptide of the invention. Kits can include any of the compositions noted above, and optionally further include additional components such as instructions to practice a high-throughput method of assaying for an effect on expression of the genes encoding a polypeptide of the invention, or on activity of a polypeptide of the invention, one or more containers or compartments (*e.g.*, to  
10 hold the probe, labels, or the like), a control modulator of the expression or activity of a polypeptide of the invention, a robotic armature for mixing kit components or the like.

[229] The invention also provides integrated systems for high-throughput screening of potential modulators for an effect on the expression or activity of a polypeptide of the invention. The systems can include a robotic armature which transfers fluid from a  
15 source to a destination, a controller which controls the robotic armature, a label detector, a data storage unit which records label detection, and an assay component such as a microtiter dish comprising a well having a reaction mixture or a substrate comprising a fixed nucleic acid or immobilization moiety.

[230] A number of robotic fluid transfer systems are available, or can easily  
20 be made from existing components. For example, a Zymate XP (Zymark Corporation; Hopkinton, MA) automated robot using a Microlab 2200 (Hamilton; Reno, NV) pipetting station can be used to transfer parallel samples to 96 well microtiter plates to set up several parallel simultaneous binding assays.

[231] Optical images viewed (and, optionally, recorded) by a camera or other  
25 recording device (*e.g.*, a photodiode and data storage device) are optionally further processed in any of the embodiments herein, *e.g.*, by digitizing the image and storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video or digitized optical image.

30 [232] One conventional system carries light from the specimen field to a cooled charge-coupled device (CCD) camera, in common use in the art. A CCD camera includes an array of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (*e.g.*, individual hybridization sites on an array of biological polymers) are sampled to obtain light intensity



readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, e.g., by fluorescent or dark field microscopic techniques.

### VIII. ADMINISTRATION AND PHARMACEUTICAL COMPOSITIONS

5                   [233] Modulators of the polypeptides of the invention (e.g., antagonists or agonists including polypeptides of the invention, fragments thereof, or fusions comprising the polypeptides or fragments which have antagonist activity or an additive effect on overall polypeptide activity) can be administered directly to the mammalian subject (typically in need thereof due to a pre-diabetic, diabetic or obese condition) for modulation of activity of a  
10 polypeptide of the invention *in vivo*. Administration is by any of the routes normally used for introducing a modulator compound into ultimate contact with the tissue to be treated and is well known to those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

15                   [234] The pharmaceutical compositions of the invention may comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there are a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g., Remington's*  
20 *Pharmaceutical Sciences*, 17<sup>th</sup> ed. 1985)).

                  [235] The modulators (e.g., agonists or antagonists) of the expression or activity of a polypeptide of the invention, alone or in combination with other suitable components, can be prepared for injection or for use in a pump device. Pump devices (also known as "insulin pumps") are commonly used to administer insulin to patients and therefore  
25 can be easily adapted to include compositions of the present invention. Manufacturers of insulin pumps include Animas, Disetronic and MiniMed.

                  [236] The modulators (e.g., agonists or antagonists) of the expression or activity of a polypeptide of the invention, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be "nebulized") to be  
30 administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

                  [237] Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers,

bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, or  
5 intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part of a prepared food or drug.

[238] The dose administered to a patient, in the context of the present  
10 invention should be sufficient to induce a beneficial response in the subject over time. The optimal dose level for any patient will depend on a variety of factors including the efficacy of the specific modulator employed, the age, body weight, physical activity, and diet of the patient, on a possible combination with other drugs, and on the severity of the case of diabetes. It is recommended that the daily dosage of the modulator be determined for each  
15 individual patient by those skilled in the art in a similar way as for known insulin compositions. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

[239] In determining the effective amount of the modulator to be  
20 administered a physician may evaluate circulating plasma levels of the modulator, modulator toxicity, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

[240] For administration, modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the  
25 modulator at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

[241] The compounds of the present invention can also be used effectively in combination with one or more additional active agents depending on the desired target therapy (see, e.g., Turner, N. et al. *Prog. Drug Res.* (1998) 51: 33-94; Haffner, S. *Diabetes Care* (1998) 21: 160-178; and DeFronzo, R. et al. (eds.), *Diabetes Reviews* (1997) Vol. 5 No. 4). A number of studies have investigated the benefits of combination therapies with oral  
30 agents (see, e.g., Mahler, R., *J. Clin. Endocrinol. Metab.* (1999) 84: 1165-71; United Kingdom Prospective Diabetes Study Group: UKPDS 28, *Diabetes Care* (1998) 21: 87-92; Bardin, C. W.,(ed.), *Current Therapy In Endocrinology And Metabolism*, 6th Edition (Mosby

- Year Book, Inc., St. Louis, MO 1997); Chiasson, J. et al., *Ann. Intern. Med.* (1994) 121: 928-935; Coniff, R. et al., *Clin. Ther.* (1997) 19: 16-26; Coniff, R. et al., *Am. J. Med.* (1995) 98: 443-451; and Iwamoto, Y. et al., *Diabet. Med.* (1996) 13 365-370; Kwiterovich, P. *Am. J. Cardiol* (1998) 82(12A): 3U-17U). These studies indicate that modulation of diabetes,  
5 among other diseases, can be further improved by the addition of a second agent to the therapeutic regimen. Combination therapy includes administration of a single pharmaceutical dosage formulation that contains a modulator of the invention and one or more additional active agents, as well as administration of a modulator and each active agent in its own separate pharmaceutical dosage formulation. For example, a modulator and a  
10 thiazolidinedione can be administered to the human subject together in a single oral dosage composition, such as a tablet or capsule, or each agent can be administered in separate oral dosage formulations. Where separate dosage formulations are used, a modulator and one or more additional active agents can be administered at essentially the same time (i.e., concurrently), or at separately staggered times (i.e., sequentially). Combination therapy is  
15 understood to include all these regimens.

[242] One example of combination therapy can be seen in treating pre-diabetic individuals (e.g., to prevent progression into type 2 diabetes) or diabetic individuals (or treating diabetes and its related symptoms, complications, and disorders), wherein the modulators can be effectively used in combination with, for example, sulfonylureas (such as  
20 chlorpropamide, tolbutamide, acetohexamide, tolazamide, glyburide, gliclazide, glynase, glimepiride, and glipizide); biguanides (such as metformin); a PPAR beta delta agonist; a ligand or agonist of PPAR gamma such as thiazolidinediones (such as ciglitazone, pioglitazone (*see, e.g.*, U.S. Patent No. 6,218,409), troglitazone, and rosiglitazone (*see, e.g.*, U.S. Patent No. 5,859,037)); PPAR alpha agonists such as clofibrate, gemfibrozil, fenofibrate,  
25 ciprofibrate, and bezafibrate; dehydroepiandrosterone (also referred to as DHEA or its conjugated sulphate ester, DHEA-SO<sub>4</sub>); antiglucocorticoids; TNF $\alpha$  inhibitors;  $\alpha$ -glucosidase inhibitors (such as acarbose, miglitol, and voglibose); amylin and amylin derivatives (such as pramlintide, (*see, also*, U.S. Patent Nos. 5,902,726; 5,124,314; 5,175,145 and 6,143,718.)); insulin secretagogues (such as repaglinide, gliquidone, and nateglinide (*see, also*, U.S. Patent  
30 Nos. 6,251,856; 6,251,865; 6,221,633; 6,174,856)), and insulin.

[243] The modulators of the invention can also be combined with anti-obesity drugs (e.g., Xenical (Orlistat), Merida (Sibutramine) or Adipex-P (Phentermine)) or appetite-suppressing drugs.

**IX. GENE THERAPY**

[244] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered amino acid sequences comprising the polypeptides of the invention in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding amino acid sequences comprising polypeptides of the invention to cells *in vitro*. In some embodiments, the nucleic acids encoding amino acid sequences comprising polypeptides of the invention are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds) (1995); and Yu *et al.*, *Gene Therapy* 1:13-26 (1994).

[245] Methods of non-viral delivery of nucleic acids encoding engineered polypeptides of the invention include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* administration).

[246] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291-297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); Gao *et al.*, *Gene Therapy* 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[247] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered polypeptides of the invention take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they  
5 can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of polypeptides of the invention could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the  
10 retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[248] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors  
15 are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used  
20 to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (*see, e.g.*, Buchscher *et al.*, *J. Virol.* 66:2731-2739 (1992); Johann *et al.*, *J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al.*,  
25 *Virol.* 176:58-59 (1990); Wilson *et al.*, *J. Virol.* 63:2374-2378 (1989); Miller *et al.*, *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[249] In applications where transient expression of the polypeptides of the invention is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell  
30 division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (*see, e.g.*, West *et al.*, *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO

93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994)). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989).

[250] pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar *et al.*, *Blood* 85:3048-305 (1995); Kohn *et al.*, *Nat. Med.* 1:1017-102 (1995); Malech *et al.*, *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al.*, *Science* 270:475-480. (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem *et al.*, *Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al.*, *Hum. Gene Ther.* 1:111-2 (1997).

[251] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al.*, *Lancet* 351:9117 1702-3 (1998), Kearns *et al.*, *Gene Ther.* 9:748-55 (1996)).

[252] Replication-deficient recombinant adenoviral vectors (Ad) can be engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiply types of tissues *in vivo*, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al.*, *Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al.*, *Infection* 24:1 5-10 (1996); Sterman *et al.*, *Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al.*, *Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al.*, *Hum. Gene Ther.* 5:597-613 (1997); Topf *et al.*, *Gene Ther.* 5:507-513 (1998); Sterman *et al.*, *Hum. Gene Ther.* 7:1083-1089 (1998).

[253] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and  $\psi$ 2 cells or

PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[254] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, *PNAS* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

[255] Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal

donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[256] *Ex vivo* cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In some embodiments, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA) encoding a polypeptides of the invention, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (*see, e.g., Freshney et al., Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

[257] In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* into clinically important immune cell types using cytokines such as GM-CSF, IFN- $\gamma$  and TNF- $\alpha$  are known (*see Inaba et al., J. Exp. Med.* 176:1693-1702 (1992)).

[258] Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (*see Inaba et al., J. Exp. Med.* 176:1693-1702 (1992)).

[259] Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[260] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention, as described below (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989).



**X. DIAGNOSIS OF OBESITY AND/OR DIABETES**

[261] The present invention also provides methods of diagnosing diabetes or obesity, or a predisposition of at least some of the pathologies of diabetes and/or obesity. Diagnosis can involve determination of a genotype of an individual (e.g., with SNPs) and comparison of the genotype with alleles known to have an association with the occurrence of obesity and/or diabetes. Alternatively, diagnosis also involves determining the level of a polypeptide or polynucleotide of the invention in a patient and then comparing the level to a baseline or range. Typically, the baseline value is representative of a polypeptide or polynucleotide of the invention in a healthy (e.g., lean) person.

[262] As discussed above, variation of levels (e.g., low or high levels) of a polypeptide or polynucleotide of the invention compared to the baseline range indicates that the patient is either obese, at risk for becoming obese, diabetic or at risk of developing at least some of the pathologies of diabetes (e.g., pre-diabetic). The level of a polypeptide in a lean individual can be a reading from a single individual, but is typically a statistically relevant average from a group of lean individuals. The level of a polypeptide in a lean individual can be represented by a value, for example in a computer program.

[263] In some embodiments, the level of polypeptide or polynucleotide of the invention is measured by taking a blood, urine or tissue sample from a patient and measuring the amount of a polypeptide or polynucleotide of the invention in the sample using any number of detection methods, such as those discussed herein. For instance, fasting and fed blood or urine levels can be tested.

[264] In some embodiments, the baseline level and the level in a lean sample from an individual, or at least two samples from the same individual differ by at least about 5%, 10%, 20%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500%, 1000% or more. In some embodiments, the sample from the individual is greater by at least one of the above-listed percentages relative to the baseline level. In some embodiments, the sample from the individual is lower by at least one of the above-listed percentages relative to the baseline level.

[265] In some embodiments, the level of a polypeptide or polynucleotide of the invention is used to monitor the effectiveness of either anti-obese therapies such as orlistat or sibutramine, or, antidiabetic therapies such as thiazolidinediones, metformin, sulfonylureas and other standard therapies. In some embodiments the activity or expression of a polypeptide or polynucleotide of the invention will be measured prior to and after treatment of an obese patient with antiobese therapies, or, diabetic or pre-diabetic patients

with antidiabetic therapies as a surrogate marker of clinical effectiveness. For example, the greater the reduction in expression or activity of a polypeptide of the invention indicates greater effectiveness.

5 [266] Glucose/insulin tolerance tests can also be used to detect the effect of glucose levels on levels of a polypeptide or polynucleotide of the invention. In glucose tolerance tests, the patient's ability to tolerate a standard oral glucose load is evaluated by assessing serum and urine specimens for glucose levels. Blood samples are taken before the glucose is ingested, glucose is given by mouth, and blood or urine glucose levels are tested at set intervals after glucose ingestion. Similarly, meal tolerance tests can also be used to detect  
10 the effect of insulin or food, respectively, on levels of a polypeptide or polynucleotide of the invention.

[267] Body weight or other indicators of obesity can also be used to detect the effect of modulating the levels of a polypeptide or polynucleotide of the invention. Measurement of a subject's response can be evaluated by assessing serum for altered levels of  
15 obesity-associated gene products, e.g., leptin, TNF alpha or IL-6.

[268] All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

20 [269] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

## 25 EXAMPLES

[270] The following examples are offered to illustrate, but not to limit the claimed invention.

30 [271] In either obese insulin-resistant or type II diabetics, peripheral tissues, especially muscle and fat, are known to have an impaired ability to respond to insulin and hence to take up glucose. This defect in glucose metabolism is usually compensated for by increased secretion of insulin from the pancreas, thereby maintaining normal glucose levels. The majority of glucose disposal occurs in the muscle. A number of obese insulin resistant patients will progress to overt diabetics with time. The molecular defects underlying this peripheral insulin resistance in both the obese and type II diabetics are not well defined.

Genes in muscle or fat whose expression is altered in either or both the obese or type II diabetics when compared to lean individuals can be causative genes for either obesity, insulin resistance and/or diabetes and are able to predict the transition to diabetes. Modulators of such genes have the ability to reverse obesity, insulin resistance and restore normal insulin sensitivity, thereby improving whole body glucose homeostasis including for example insulin secretion. Modulators of such genes also have the ability to be used to pre-empt the transition from obesity-induced insulin resistance to diabetes. Modulators of such genes also have the ability to be used to reverse metabolic obesity-related diseases such as cardiovascular disease, hypertension or obesity-related cancer.

[272] The molecular mechanism by which thiazolidinediones (TZDs) cause an increase in peripheral insulin sensitivity was studied. Genes in muscle or fat whose expression is altered by TZDs may lie on a pathway leading from TZD treatment to increased insulin sensitivity. Modulators of such genes can elicit the same effect as TZD treatment. Moreover, such modulators can lack some of the side effects of TZD. Gene expression profiling in cultures of primary human adipocytes treated with either pioglitazone or rosiglitazone were used to identify genes important for TZD action and therefore treatment of obesity, diabetes and/or insulin resistance.

[273] Gene expression profiling was performed on tissue samples (subcutaneous adipose samples) obtained from lean, obese and diabetic individuals. Two studies were performed. In the first study, samples were isolated from all individuals after a 5 hour hyperinsulinemic euglycemic clamp.

[274] In the second study, subcutaneous adipose samples were obtained from lean (BMI < 25) and obese (BMI > 30) individuals after an overnight fast.

[275] In a third study samples were obtained from human subcutaneous and omental adipose tissues. Genes expressed only, or enriched, in fat can lie on pathways involved in insulin sensitivity, appetite suppression or lipid metabolism in the adipose itself or other peripheral tissues (e.g., muscle, liver, brain). For all tissue samples mRNA was isolated from these adipose samples and converted to cRNA by standard procedures. The gene expression profile for each individual was determined by hybridization of cRNA to commercial and custom synthesized Affymetrix chips.

[276] Gene expression profile differences were calculated as follows. The expression level of a particular gene is indicated by its 'signal intensity'. The raw data was analyzed by a statistical test to remove 'outliers'. The mean 'signal intensity' was then calculated from the signal intensities for all individuals in a particular treatment group.

Genes were determined to be changed in the first two studies by calculating the Students t test statistic between the two conditions and selecting those with t less than or equal to 0.05. The fold change was determined as the ratio of mean signal intensity in condition 2 to the mean signal intensity in condition 1. In the first study three comparisons was undertaken: diabetics (condition 1) versus leans (condition 2), obese (condition 1) versus lean (condition 2) and diabetics (condition 1) versus obese (condition 2). The second study comparison is lean (condition 1) versus obese (condition 2). The third comparison is identification of fat specific or fat enriched genes when comparing the expression profile of human subcutaneous and omental adipose tissues to at last 12 other human adult tissues. Genes were determined to be meeting the criteria cut-off when the mean signal intensity of the human adipose samples was 3 fold greater than the mean signal intensity of all the other human adult tissues profiled or called present only in the adipose samples and absent in all others by the Affymetrix software program.

#### ADLICAN

[277] Probe set 209596 detects ADLICAN nucleic acid sequences.

Expression of ADLICAN transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	126.4	17.87	10	80.24	8.1	8	1.58	0.037	ADLICAN

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

Change" indicates fold change of diabetics in comparison to lean patients.

[278] ADLICAN was also evaluated using real-time PCR. The results further show that ADLICAN is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	2.04	0.115

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic

expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[279] Probe set 209596 detects ADL1CAN nucleic acid sequences.

Expression of ADL1CAN transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	195.64	30.94	5	94.43	8.71	4	2.07	0.028	ADL1CAN

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 5 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[280] ADL1CAN was also evaluated using real-time PCR. The results further show that ADL1CAN is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5)/ Lean (4)	4.35	0.073

- 10 "Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[281] Probe set 209596 detects ADL1CAN nucleic acid sequences.

- 15 Expression of ADL1CAN transcripts was decreased in pio compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	PIO			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	31.8	3.53	3	51.9	4.05	3	0.61	0.021	ADL1CAN

B/C indicates sample is from Basal or Clamp; "Pre-Pio" and "Post-Pio" indicates sample was

taken before or after 24 hours of pioglitazone treatment; "Mean Expr" indicates mean

expression; "SEM" indicates standard error of mean; "n" indicates number of patient

samples; "Fold Change" indicates fold change of primary human adipocytes post-pio in

- 20 comparison to pre-pio samples.

[282] Probe set 209596 detects ADL1CAN nucleic acid sequences.

Expression of ADL1CAN transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
142.7	13.24	5	37.71	3.55	13	3.78	0.001	ADL1CAN

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [283] ADL1CAN contains the following protein domains (designated with reference to SEQ ID NO:2): Atrophin-1 family (PF03154) at amino acids 1405 to 2232; Leucine rich repeat N-terminal domain (PF01462) at amino acids 26 to 54; Geminivirus AL2 protein (PF01440) at amino acids 1317 to 1428; Leucine Rich Repeat (PF00560) at amino acids 80 to 103, 128 to 151; and Immunoglobulin domain (PF00047) at amino acids 494 to 557, 592 to 653, 1868 to 1930, 1965 to 2027, 2062 to 2124, 2161 to 2223, 2258 to 2326, 2361 to 2420, 2459 to 2520, 2557 to 2618, 2652 to 2713, 2748 to 2812. ADL1CAN is a protein which contains many domains which mediate protein-protein binding.

### ALDH1A3

[284] Probe set 203180 detects ALDH1A3 nucleic acid sequences.

- Expression of ALDH1A3 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	58.99	6.02	10	26.64	3.98	8	2.21	<0.000	ALDH1A3

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [285] ALDH1A3 was also evaluated using real-time PCR. The results further show that ALDH1A3 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	2.21	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[286] Probe set 203180 detects ALDH1A3 nucleic acid sequences.

Expression of ALDH1A3 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	42.3	3.9	8	26.64	3.98	8	1.59	0.018	ALDH1A3

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 5 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[287] ALDH1A3 was also evaluated using real-time PCR. The results further show that ALDH1A3 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (8)/ Lean (8)	1.59	0.018

- 10 "Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[288] Probe set 203180 detects ALDH1A3 nucleic acid sequences.

- 15 Expression of ALDH1A3 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	42.76	7.99	5	18.65	1.47	4	2.29	0.038	ALDH1A3

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- 20 [289] ALDH1A3 was over-expressed in L6 myotubes and the effect on basal and insulin stimulated glucose transport was determined

#### Glucose Transport Analysis in L6 Myotubes

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC.. (ALDH1A3/Con) +/- SEM (n=3)
	FC (hALDH1 A3/Con)	t test ( $\alpha$ )	FC (hALDH1 A3/Con)	t test ( $\alpha$ )	FC (hALDH1 A3/Con)	t test ( $\alpha$ )	
0	1.27	0.001	1.22	0.001	1.16	0.001	1.22±0.03
10	1.29	0.001	1.07	0.532	1.11	0.050	1.16±0.07

100	1.29	0.001	1.14	0.057	1.03	0.634	1.15±0.08
-----	------	-------	------	-------	------	-------	-----------

**Legend** "Con" indicates control L6 myotubes that do not express hALDH1A3. "FC"

indicates the fold change defined as the following ratio; glucose transport in hALDH1A3-expressing cells/glucose transport in non-hALDH1A3-expressing cells. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

5                    [290] The results show that increased levels of ALDH1A3 in a cell such as a muscle cell leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of ALDH1A3 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

10                   [291] ALDH1A3 contains the following protein domains (designated with reference to SEQ ID NO:8): Aldehyde dehydrogenase family (PF00171) at amino acids 40 to 507. ALDH1A3 is a retinaldehyde dehydrogenase that catalyzes the oxidation of all-trans-retinaldehyde to retinoic acid and may have a role in cell differentiation and proliferation (Grun, F., *et al J Biol Chem.* 275: 41210-8 (2000); Rexer, B. N., *et al Cancer Res.* 61: 7065-15 7070 (2001).

                    [292] It has been established that the mRNA for ALDH1A3 can be induced in hepatocytes by agents such as omeprazole (Nishimura *et al Yakugaku Zasshi.* 122 :339-61 (2002)). Thus, an exemplary method in which ALDH1A3 activators can be identified comprises treating hepatocytes with candidate compounds and measuring increases in 20 ALDH1A3 mRNA.

### ALK7

                    [293] Probe set MBXHUMFAT04495 detects ALK7 nucleic acid sequences. Expression of ALK7 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	173.96	12.02	10	241.5	20.33	8	0.72	0.015	ALK7

25 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.



[294] ALK7 was also evaluated using real-time PCR. The results further show that ALK7 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.65	0.017

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic

- 5 expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[295] Probe set MBXHUMFAT04495 detects ALK7 nucleic acid sequences. Expression of ALK7 transcripts was decreased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

CORRELATION RD				
B/C	n	Corr Co-efficient	Students t test	Gene Name
C	18	0.623	<0.005	ALK7

- 10 B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

- [296] ALK7 was also evaluated using real-time PCR. The results further  
15 show that ALK7 is significantly decreased in patients with insulin resistance compared to normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (18)	0.534	<0.02

"Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number  
20 of patient samples analyzed by real-time PCR.

[297] Probe set MBXHUMFAT04495 detects ALK7 nucleic acid sequences. Expression of ALK7 transcripts was decreased in obese compared to lean patients in the gene profiling experiment.

OBESE				LEAN					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
B	160.52	14.73	5	313.3	28.95	4	0.51	0.007	ALK7

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 25 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[298] Probe set MBXHUMFAT04495 detects ALK7 nucleic acid sequences.

Expression of ALK7 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
158.78	12.8	5	15.65	4.07	13	10.15	<0.000	ALK7

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

- 5 indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

[299] ALK7 was over-expressed in L6 myotubes and the effect on basal and insulin stimulated glucose transport was determined

#### 10 Glucose Transport Analysis in L6 Myotubes

	Experiment 1		Experiment 2		Experiment 3		
Insulin (nM)	FC (ALK7 /Con)	t test ( $\alpha$ )	FC (ALK7 /Con)	t test ( $\alpha$ )	FC (ALK7 /Con)	t test ( $\alpha$ )	Mean FC (ALK7/Con) +/- SD (n=3)
0	0.85	0.010	0.86	0.049	0.86	0.017	0.85±0.01
10	0.83	0.001	0.84	0.001	0.89	0.002	0.85±0.03
100	0.84	0.002	0.94	0.251	0.89	0.048	0.89±0.05

Legend "Con" indicates control L6 myotubes that do not express hALK7. "FC" indicates the fold change defined as the following ratio; glucose transport in hALK7-expressing L6 myotubes/glucose transport in non-ALK7-expressing L6 myotubes. "h" is human. "n" is the number of experiments. SD is the standard deviation.

15

[300] The results show that increased levels of ALK7 in a cell such as a muscle cell leads to a corresponding decrease in glucose uptake. This indicates that decreasing the levels or activity of ALK7 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

20

[301] ALK7 contains the following protein domains (designated with reference to SEQ ID NO:14): Signal peptide at amino acids 1 to 25; Activin types I and II receptor domain (PF01064) at amino acids 15 to 100; Protein kinase domain (PF00069) at amino acids 195 to 482; u-PAR/Ly-6 domain (PF00021) at amino acids 30 to 94; and 1 transmembrane domain (TMHMM2.0) at amino acids 114 to 136. ALK7 is a transmembrane receptor protein serine-threonine kinase for the transforming growth factor-beta (TGF-beta) superfamily related growth factors and signals through SMAD2 (Bondestam J. *et al.*,

25

*Cytogenet. Cell Genet.*, 95: 157-162 (2001). Nodal was identified as the ligand for ALK7. ALK7 may play a role in proliferation and apoptosis (Munir, S., *et al.*, *J Biol Chem.* May 18 Epub (2004); Jornvall, H., *et al.*, *J Biol Chem.* 276: 5140-6. (2001)).

- [302] ALK7 is a type I serine/threonine kinase receptor of the transforming growth factor (TGF)-beta family. Signalling from the ALK7 receptor involves phosphorylation of SMAD2 and SMAD3 (*see, e.g.*, Kim J, *et al.*, *J. Biol. Chem* 279: 28458-28465 (2004)) Inhibitors of ALK7 kinase activity can thus be identified, for example, by using an *in vitro* phosphorylation assay containing recombinant ALK7 incubated with recombinant SMAD2 or SMAD3 and radio-labelled ATP
- [303] Inhibitors of ALK7 kinase activity such as SB 505124 (2-(5-benzol[1,3] dioxol-5-yl-2-tert-butyl-3H-imidazol-4-yl)-6-methylpyridine hydrochloride (*e.g.*, Byfield *et al.*, *Mol. Pharmacol* 65 744-752 (2004) and SB 431542 (4-(5-benzol[1,3] dioxol-5-yl-4-pyridin-2-yl-1H-imidazol-2-yl)-benzamide (Inman *et al.*, *Mol Pharmacol* 62 65-74 (2002)) are known. Such inhibitors as well as other ALK7 kinase inhibitors, *e.g.*, identified using screening assays as described herein can be used to treat insulin resistance and diabetes.

### C3AR1

[304] Probe set 209906 detects C3AR1 nucleic acid sequences. Expression of C3AR1 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	98.52	6.15	10	55.65	13.05	8	1.77	0.014	C3AR1

- B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [305] C3AR1 was also evaluated using real-time PCR. The results further show that C3AR1 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.89	0.008

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[306] Probe set 209906 detects C3AR1 nucleic acid sequences. Expression of C3AR1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	84.24	8.87	5	33.8	3.56	4	2.49	0.003	C3AR1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 5 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[307] The cellular level of C3AR1 was reduced in 3T3-L1 adipocytes using siRNA directed against C3AR1 and the effect on basal and insulin stimulated glucose transport was determined

10

C3AR1 mRNA Level in 3T3-L1 Adipocytes Transfected with siRNA Oligonucleotides

	Expt 1	Expt 2	Expt 3	Expt 4	Mean FC ± SEM (n=4)
FC (siRNA/Scr)	0.53	0.31	0.39	0.35	0.40 ± 0.05

- 15 **Legend:** "siRNA" indicates Dharmacon Smartpool siRNA oligonucleotides directed against murine C3AR1. "Scr" indicates the Dharmacon Scramble siRNA Control. "FC" indicates the fold change defined as the following ratio; Level of C3AR1 mRNA in C3AR1 siRNA transfected 3T3-L1 adipocytes/Level of C3AR1 mRNA in Scramble siRNA transfected 3T3-L1 adipocytes. "n" is the number of experiments. SEM is the standard error of the mean.

Glucose Transport in 3T3-L1 Adipocytes Transfected with siRNA Oligonucleotides

	Expt 1	Expt 2	Expt 3	Expt 4	Mean FC ± SEM (n=4)	Mean t-test
Insulin (nM)	FC (siRNA/Scr)	FC (siRNA/Scr)	FC (siRNA/Scr)	FC (siRNA/Scr)	FC (siRNA/Scr)	FC (siRNA/Scr)
0.0	0.602	0.708	0.673	0.531	0.661 ± 0.03	0.043
0.1	0.545	0.412	0.452	0.399	0.469 ± 0.03	0.004
1.0	0.588	0.715	0.553	0.546	0.619 ± 0.04	0.035
10.0	0.660	0.858	0.601	0.563	0.706 ± 0.07	0.084

- 20 **Legend:** "siRNA" indicates Dharmacon Smartpool siRNA oligonucleotides directed against murine C3AR1. "Scr" indicates Dharmacon Scramble siRNA Control oligonucleotides. "FC" indicates the fold change defined as the following ratio; glucose transport in C3AR1 siRNA transfected 3T3-L1 adipocytes/glucose transport in Scramble

siRNA transfected 3T3-L1 adipocytes. "n" is the number of experiments. SEM is the standard error of the mean.

[308] These results show that decreasing the levels of C3AR1 in a cell such as an adipocyte leads to a corresponding decrease in glucose uptake. This indicates that increasing the levels or activity of C3AR1 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[309] C3AR1 contains the following protein domains (designated with reference to SEQ ID NO:20): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 40 to 435; and 7 transmembrane domains (TMHMM2.0) at amino acids 24 to 46, 59 to 81, 96 to 118, 138 to 160, 338 to 360, 380 to 402, 417 to 439. C3AR1 is the G protein-coupled receptor for complement component 3a and mediates various aspects of inflammatory responses including complement activation and chemotaxis (Fischer, W. H and Hugli T.E., *J. Immunol.* 159: 4279-4286 (1997); Zwirner, J., *et al.*, *Eur J Immunol* 28: 1570-7. (1998); Crass, T., *et al.*, *Eur J Immunol* 26: 1944-1950 (1996)).

[310] C3AR1 is a G protein coupled receptor, activation of which results in the release of intracellular Ca<sup>2+</sup> in HMC-1 cells (*see, e.g.*, Legler, D.F. *et al.*, *Eur.J.Immunol* 26: 753-758 (1996)). Agonists of the C3AR1 can therefore be identified, for example, using assays that measure changes in intracellular calcium. An exemplary assay is a cell based assay in which cells over-expressing C3AR1, such as HMC-1 cells, are treated with compounds and an increase in intracellular Ca<sup>2+</sup> is measured using Ca<sup>2+</sup> sensitive dyes such as Calcium 3.

### CALCRL

[311] Probe set 210815 detects CALCRL nucleic acid sequences. Expression of CALCRL transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	56.16	4.08	10	85.81	5.59	8	0.65	0.001	CALCRL

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

Change" indicates fold change of diabetics in comparison to lean patients.

[312] CALCRL was also evaluated using real-time PCR. The results further show that CALCRL is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.65	0.001

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic

5 expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[313] CALCRL was over-expressed in 3T3-L1 adipocytes and the effect on basal and insulin stimulated glucose transport was determined.

Glucose Transport Analysis in 3T3-L1 Adipocytes:

	Experiment 1		Experiment 2		Experiment 3		
Insulin (nM)	FC (hCALCR L / Con)	t test ( $\alpha$ )	FC (hCALCR L / Con)	t test ( $\alpha$ )	FC (hCALCR L / Con)	t test ( $\alpha$ )	Mean FC (hCALCR/Con) +/- SEM (n=3)
0	1.22	0.047	1.09	0.216	1.31	0.045	1.21±0.07
0.05	1.01	0.895	1.05	0.454	1.12	0.609	1.06±0.03
0.1	1.03	0.508	1.05	0.513	1.22	0.077	1.10±0.06
0.3	0.99	0.862	1.01	0.805	1.16	0.572	1.05±0.05
1	1.08	0.007	1.01	0.774	1.13	0.604	1.07±0.03
10	1.05	0.300	1.05	0.068	1.06	0.584	1.06±0.001

10 **Legend:** "Con" indicates control 3T3-L1 adipocytes that do not express hCALCR. "FC" indicates the fold change defined as the following ratio; glucose transport in hCALCR-expressing 3T3-L1 adipocytes /glucose transport in non-hCALCR expressing 3T3-L1 adipocytes. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

15

[314] The results show that increasing the levels of CALCRL in a cell such as an adipocyte leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of CALCRL in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[315] CALCRL contains the following protein domains (designated with reference to SEQ ID NO:26): Signal peptide at amino acids 1 to 22; Hormone receptor domain (PF02793) at amino acids 62 to 132; 7 transmembrane receptor (Secretin family) (PF00002) at amino acids 138 to 391; and 7 transmembrane domains (TMHMM2.0) at amino acids 144 to 166, 179 to 198, 225 to 247, 254 to 276, 291 to 313, 334 to 352, 367 to 389. CALCRL is the G protein-coupled receptor which binds calcitonin-gene-related peptide

(CGRP) or adrenomedullin (ADM) depending upon interaction with either of the accessory proteins, RAMP1 and RAMP2 and stimulates adenylyl cyclase (Kamitani, S., *et al.*, *FEBS Lett.* 448: 111-114 (1999); Kuwasako, K., *et al.*, *Mol Pharmacol.* 65: 207-13 (2004); Flahaut, M., *et al.*, *Biochemistry.* 42: 10333-41 (2003)).

## 5 CCL13

[316] Probe set 206407 detects CCL13 nucleic acid sequences. Expression of CCL13 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	43.25	3.49	10	29.01	4.63	8	1.49	0.028	CCL13

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 10 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[317] Probe set 206407 detects CCL13 nucleic acid sequences. Expression of CCL13 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	34.1	1.97	5	15.77	0.18	4	2.16	0.003	CCL13

- 15 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [318] CCL13 was also evaluated using real-time PCR. The results further show that CCL13 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.
- 20

Comparison	Expression Fold Change	t test
Obese (5)/ Lean (4)	2.88	0.048

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [319] CCL13 contains the following protein domains (designated with reference to SEQ ID NO:32): Signal peptide at amino acids 1 to 23; and Small cytokines (intecrine/chemokine), interleukin-8 like (PF00048) at amino acids 24 to 89. A soluble active
- 25

secreted form of CCL13 has been detected (Berkhout, T.A., *et al.*, *J Biol Chem.* 272:16404-13 (1997)) and this is displayed in SEQ ID NO:33. CCL13 displays chemotactic activity for monocytes, lymphocytes, basophils and eosinophils, but not neutrophils. This chemokine plays a role in accumulation of leukocytes during inflammation. It may also be involved in the recruitment of monocytes into the arterial wall during atherosclerosis (Garcia-Zepeda, E. A. *et al.*, *J Immunol* 157: 5613-5626 (1996); White, J. R. *et al.*, *J Biol Chem* 275: 36626-36631 (2000); Wain, J.H., *et al.*, *Clin Exp Immunol.* 127: 436-44 (2002)).

### CCL8

[320] Probe set 214038 detects CCL8 nucleic acid sequences. Expression of CCL8 transcripts was decreased in pio compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	PIO			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	7.37	2.62	3	21.87	0.82	3	0.34	0.023	CCL8

B/C indicates sample is from Basal or Clamp; "Pre-Pio" and "Post-Pio" indicates sample was taken before or after 24 hours of pioglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-pio in comparison to pre-pio samples.

[321] CCL8 was also evaluated using real-time PCR. The results further show that CCL8 is significantly under-expressed in primary cultured human adipocytes treated with pio when compared to vehicle.

Comparison	Expression Fold Change	t test
Post-Pio (12) / Pre-Pio (12)	0.3	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean pio expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

[322] Probe set 214038 detects CCL8 nucleic acid sequences. Expression of CCL8 transcripts was decreased in rosi compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.



B/C	ROSI			VEHICLE					
	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
B	10.3	1.63	3	21.87	0.82	3	0.47	0.008	CCL8

B/C indicates sample is from Basal or Clamp; "Pre-Rosi" and "Post-Rosi" indicates sample was taken before or after 24 hours of rosiglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-rosi in comparison to pre-rosi samples.

[323] CCL8 was also evaluated using real-time PCR. The results further show that CCL8 is significantly under-expressed in primary cultured human adipocytes treated with rosi when compared to vehicle.

Comparison	Expression Fold Change	t test
Post-Rosi (12)/ Pre-Rosi (12)	0.42	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean rosi expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

[324] Probe set 214038 detects CCL8 nucleic acid sequences. Expression of CCL8 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
62.4	21.71	5	11.92	5.23	13	5.24	0.08	CCL8

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

[325] CCL8 contains the following protein domains (designated with reference to SEQ ID NO:39): Signal peptide at amino acids 1 to 23; and Small cytokines (intecrine/chemokine), interleukin-8 like (PF00048) at amino acids 24 to 90. A soluble active secreted form of CCL8 has been detected (Van Damme, J. *et al.*, *J Exp Med.* 176: :59-65 (1992)) and this is displayed in SEQ ID NO:40. CCL8 displays chemotactic activity for monocytes, lymphocytes, basophils and eosinophils. By recruiting leukocytes to sites of inflammation this cytokine may contribute to tumor-associated leukocyte infiltration and to the antiviral state against HIV infection ( Noso, N. *et al.*, *Biochem. Biophys. Res. Commun.* 200: 1470-1476 (1994); Yang, O. *et al.*, *J. Infect. Dis.* 185: 1174-1178 (2002)).

**CHI3L1**

[326] Probe set 209395 detects CHI3L1 nucleic acid sequences. Expression of CHI3L1 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	84.62	18.54	10	24.83	9.44	8	3.41	0.013	CHI3L1

- 5 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [327] CHI3L1 was also evaluated using real-time PCR. The results further show that CHI3L1 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	3.38	0.037

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [328] Probe set 209395 detects CHI3L1 nucleic acid sequences. Expression of CHI3L1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	39.26	8.12	5	12.25	6.2	4	3.2	0.034	CHI3L1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [329] Probe set 209395 detects CHI3L1 nucleic acid sequences. Expression of CHI3L1 transcripts was decreased in pio compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	PIO			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	109.03	15.04	3	257.9	30.35	3	0.42	0.023	CHI3L1

B/C indicates sample is from Basal or Clamp; "Pre-Pio" and "Post-Pio" indicates sample was taken before or after 24 hours of pioglitazone treatment; "Mean Expr" indicates mean

expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-pio in comparison to pre-pio samples.

- [330] CHI3L1 was also evaluated using real-time PCR. The results further show that CHI3L1 is significantly under-expressed in primary cultured human adipocytes treated with pio when compared to vehicle.

Comparison	Expression Fold Change	t test
Post-Pio (12)/ Pre-Pio (12)	0.22	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean pio expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

- [331] Probe set 209395 detects CHI3L1 nucleic acid sequences. Expression of CHI3L1 transcripts was decreased in rosi compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	ROSI			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	102.77	16.32	3	257.9	30.35	3	0.4	0.019	CHI3L1

- B/C indicates sample is from Basal or Clamp; "Pre-Rosi" and "Post-Rosi" indicates sample was taken before or after 24 hours of rosiglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-rosi in comparison to pre-rosi samples.

- [332] CHI3L1 was also evaluated using real-time PCR. The results further show that CHI3L1 is significantly under-expressed in primary cultured human adipocytes treated with rosi when compared to vehicle.

Comparison	Expression Fold Change	t test
Post-Rosi (12)/ Pre-Rosi (12)	0.25	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean rosi expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

- [333] CHI3L1 was over-expressed in L6 myotubes and the effect on basal and insulin stimulated glucose transport was determined.

Glucose Transport Analysis in L6 Myotubes

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (CHI3L1/Con) +/- SEM (n=3)
	FC (CHI3L1/ Con)	t test ( $\alpha$ )	FC (CHI3L1/ Con)	t test ( $\alpha$ )	FC (CHI3L1/ Con)	t test ( $\alpha$ )	
0	1.07	0.277	1.18	0.451	1.17	0.040	1.14±0.03
10	1.05	0.034	1.32	0.024	1.18	0.048	1.18±0.08
100	1.11	0.103	1.37	0.026	1.24	0.026	1.24±0.07

**Legend** "Con" indicates control L6 myotubes that do not express CHI3L1. "FC" indicates the fold change defined as the following ratio; glucose transport in CHI3L1-expressing cells/glucose transport in non-CHI3L1-expressing cells. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

[334] The results show that increasing the levels of CHI3L1 in a cell such as a muscle cells leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of CHI3L1 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[335] CHI3L1 contains the following protein domains (designated with reference to SEQ ID NO:46): Signal peptide at amino acids 1 to 21; and Glycosyl hydrolases family 18 (PF00704) at amino acids 22 to 357. A soluble active secreted form of CHI3L1 has been detected (Hakala, B.E., *et al.*, *J Biol Chem.* 268:25803-10 (1993)) and this is displayed in SEQ ID NO:47. CHI3L1 is a glycoprotein secreted by a variety of cells including articular chondrocytes, synoviocytes and macrophages (Recklies, A.D., *et al.*, *Biochem J.* 365: 119-26 (2002)) and is associated with conditions of increased matrix turnover and tissue remodeling for example, arthritis (Punzi, L., *et al.*, *Ann Rheum Dis.* 62: 1224-6 (2003); Ling, H. and Recklies, A.D., *Biochem J.* Mar 12;Pt. Epub (2004)).

[336] The CHI3L1 gene has been cloned and the proximal promoter has been identified and shown to contain binding sites for transcription factors such as PU.1, Sp1, Sp3, USF, AML-1 and C/EBP proteins (*see, e.g.*, Rehli M., *et al. Genomics.* 43: 221-225 (1997); Rehli, M., *et al. J. Biol. Chem* 278: 44058-44067 (2003)). An exemplary method of screening for CHI3L1 regulators comprises an assay as follows: A CHI3L1 promoter is inserted upstream of a reporter gene such as  $\beta$ -galactosidase and expressed in cells. Compounds that up-regulate the activity of the promoter can thus be identified by measuring increased  $\beta$ -galactosidase activity.

**CR1**

[337] Probe set 244313 detects CR1 nucleic acid sequences. Expression of CR1 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

	DIABETIC			LEAN					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
C	23.01	2.94	10	12.98	2.39	8	1.77	0.018	CR1

- 5 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [338] CR1 was also evaluated using real-time PCR. The results further show that CR1 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.96	0.004

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [339] Probe set 244313 detects CR1 nucleic acid sequences. Expression of CR1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

	OBESE			LEAN					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
B	18.95	0.87	5	10.73	1.33	4	1.77	0.003	CR1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [340] CR1 contains the following protein domains (designated with reference to SEQ ID NO:53): Sushi domain (SCR repeat) (PF00084) at amino acids 43 to 99, 104 to 161, 166 to 232, 238 to 293, 297 to 353, 358 to 416, 421 to 487, 493 to 549, 554 to 611, 616 to 682, 688 to 743, 747 to 803, 808 to 866, 871 to 937, 943 to 999, 1004 to 1061, 1066 to 1132, 1138 to 1193, 1197 to 1253, 1258 to 1316, 1321 to 1387, 1393 to 1449, 1454 to 1511, 1516 to 1582, 1588 to 1643, 1647 to 1703, 1708 to 1766, 1771 to 1837, 1846 to 1902, 1907 to 1964, 1969 to 2035, 2041 to 2096, 2100 to 2156, 2161 to 2219, 2224 to 2290, 2298 to

2354, 2359 to 2415 and 1 transmembrane domain (TMHMM2.0) at amino acids 2447 to 2489. Complement receptor 1 (CR1) is a cell surface glycoprotein on erythrocytes, leukocytes, and other cells that inhibits both the classic and alternative pathways of complement activation. with both the classical and alternative pathways. CR1 also mediates other key immunological functions such as the transport of C3b-coated immune complexes in erythrocytes, activation of phagocytosis of C3b-bearing particles by neutrophils and monocytes, induction of interleukin 1 secretion by monocytes and enhancement of B-cell differentiation (Hamer, I., *et al. Biochem. J.* 329, 183-190 (1998); Makrides, S. C. *et al. J. Biol. Chem.* 267: 24754-24761 (1992)).

## 10 CSFR1

[341] Probe set 203104 detects CSFR1 nucleic acid sequences. Expression of CSFR1 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	158.57	12.55	10	91.33	11.01	8	1.74	0.001	CSFR1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

15 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[342] CSFR1 was also evaluated using real-time PCR. The results further show that CSFR1 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.68	0.009

20 "Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[343] Probe set 203104 detects CSFR1 nucleic acid sequences. Expression of CSFR1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	148.1	8.8	5	79.28	16.17	4	1.87	0.015	CSFR1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- 5 [344] CSF1R was over-expressed in 3T3-L1 adipocytes and cells were then treated with CSF1. The effect on basal and insulin stimulated glucose transport and Glut 4 translocation was then determined

Glucose Transport Analysis in 3T3-L1 Adipocytes:

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (hCSF1R/Con) +/- SEM (n=3)
	FC (CSF1R/Con)	t test ( $\alpha$ )	FC (CSF1R/Con)	t test ( $\alpha$ )	FC (CSF1R/Con)	t test ( $\alpha$ )	
0	5.41	0.001	3.88	0.001	3.92	0.001	4.40±0.5
0.03	3.06	0.133	3.60	0.008	2.85	0.003	3.17±0.23
0.3	1.20	0.453	1.11	0.348	1.06	0.594	1.12±0.04
3	0.73	0.065	0.91	0.487	0.91	0.149	0.85±0.06

- 10 **Legend:** "Con" indicates control 3T3-L1 adipocytes that do not express CSF1R. "FC" indicates the fold change defined as the following ratio; glucose transport in hCSF1R-expressing cells stimulated with 100 ng/ml hCSF1 for 24 hours/glucose transport in non-hCSF1R-expressing cells stimulated with 100 ng/ml hCSF1 for 24 hours. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

- 15 Glut4 Translocation Analysis:

Insulin (nM)	Fold Change (Mean hCSF1R/Mean LacZ) (n=3)	t test (hCSF1R vs LacZ)
0	6.13	0.001
0.5	1.59	0.001
10	1.06	0.004

- 20 **Legend** "Fold Change" indicates the following ratio; (Mean % of hCSF1R-expressing cells incubated with 100 ng/mL murine CSF1 for 24 hours that were scored positive for cell surface Glut4)/(Mean % of LacZ-expressing cells cells incubated with 100 ng/mL murine CSF1 for 24 hours that were scored positive for cell surface Glut4). "h" is human. "n" is the number of experiments

[345] The results show that increasing the levels of CSF1R in a cell such as a adipocyte leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of CSF1R in tissues of insulin resistant patients or diabetic patients will

increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[346] CSFR1 contains the following protein domains (designated with reference to SEQ ID NO:57): Signal peptide at amino acids 1 to 19; Protein kinase domain (PF00069) at amino acids 582 to 910; two immunoglobulin domain (PF00047) at amino acids 217 to 280, 412 to 487 and 1 transmembrane domain (TMHMM2.0) at amino acids 515 to 537. CSFR1 is the tyrosine kinase receptor for colony stimulating factor 1, a cytokine which controls the production, differentiation, and function of macrophages and may be associated with advanced-stage breast carcinoma and myeloid leukemia (Sapi, E., *Exp Biol Med.* 229:1-11 (2004); Boultonwood, J. *et al.*, *Proc Natl Acad Sci U S A* 88: 6176-6180 (1991); Sapi, E. *et al.*, *Cancer Res.* 59: 5578-85 (1999); Fixe, P. and Praloran, V. *Cytokine* 10:32-7 (1998)).

[347] Cells over-expressing CSF1R can be generated (*see, e.g.*, Murray L.J., *et al. Clin Exp Metastasis.* 20: 757-66 (2003)). Agonists of the CSF1R can be identified, *e.g.*, by screening such cells for compounds that have the ability to induce autophosphorylation of the CSF1R.

## **CTSK**

[348] Probe set 202450 detects CTSK nucleic acid sequences. Expression of CTSK transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	227.26	9.47	10	176.54	9.77	8	1.29	0.002	CTSK

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[349] CTSK was also evaluated using real-time PCR. The results further show that CTSK is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.68	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.



[350] Probe set 202450 detects CTSK nucleic acid sequences. Expression of CTSK transcripts was increased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

CORRELATION RD				
B/C	n	Corr Co-efficient	Students t test	Gene Name
C	26	-0.689	<0.005	CTSK

B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship

- 5 between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

- [351] CTSK was also evaluated using real-time PCR. The results further show that CTSK is significantly increased in patients with insulin resistance compared to  
10 normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (26)	-0.74	<0.005

"Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- 15 [352] Probe set 202450 detects CTSK nucleic acid sequences. Expression of CTSK transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
226.1	32.74	5	62.38	10.98	13	3.62	0.005	CTSK

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in  
20 comparison to all other human adult tissues profiled.

- [353] CTSK contains the following protein domains (designated with reference to SEQ ID NO:63): Signal peptide at amino acids 1 to 23; Outer membrane lipoprotein LolB (PF03550) at amino acids 4 to 158; and Papain family cysteine protease (PF00112) at amino acids 115 to 328. A soluble active secreted form of CTSK has been  
25 detected and this is displayed in SEQ ID NO:64. CTSK is a cysteine (thiol) protease involved in bone remodeling and reabsorption, acts as a collagenase towards cartilage proteoglycans and may play a role in extracellular matrix degradation. Mutations in this gene are the cause

of pycnodysostosis, an autosomal recessive disease characterized by osteosclerosis and short stature. (Motyckova, G. and Fisher, D.E., *Curr Mol Med.* 2: 407-21(2002); Soderstrom, M. *et al.*, *Biochim Biophys Acta* 1446: 35-46 (1999); Hou, W.S., *et al.*, *Biol Chem.* 384: 891-7 (2003)).

## 5 CXCR4

[354] Probe set 211919 detects CXCR4 nucleic acid sequences. Expression of CXCR4 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	60.88	5.37	5	37.45	5.25	4	1.63	0.027	CXCR4

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 10 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[355] CXCR4 was also evaluated using real-time PCR. The results further show that CXCR4 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	2.79	0.001

- 15 "Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR

- 20 [356] CXCR4 was over-expressed in 3T3-L1 adipocytes and the cells were treated with SDF1, a ligand for CXCR4. The effects on basal and insulin stimulated glucose transport and Glut 4 translocation were determined.

### Glucose Transport Analysis in 3T3-L1 Adipocytes:

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (hCXCR4/Con) +/- SEM (n=3)
	FC (hCXCR4/Con)	t test ( $\alpha$ )	FC (hCXCR4/Con)	t test ( $\alpha$ )	FC (hCXCR4/Con)	t test ( $\alpha$ )	
0	2.35	0.001	1.54	0.001	0.95	0.48	1.61±0.40
0.03	1.64	0.014	1.62	0.004	0.71	0.07	1.32±0.31
0.3	1.79	0.003	1.05	0.341	0.96	0.73	1.27±0.26
3	1.71	0.006	0.99	0.944	1.04	0.77	1.25±0.23

**Legend:** "Con" indicates control 3T3-L1 adipocytes that do not express hCXCR4. "FC" indicates the fold change defined as the following ratio; glucose transport in hCXCR4-expressing cells incubated for 24 hours with 20 nM SDF1/glucose transport in non-CXCR4-expressing cells incubated for 24 hours with 20 nM SDF1. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

**Glut4 Translocation Analysis:**

Insulin (nM)	Fold Change (Mean hCXCR4/Mean LacZ) (n=3)	t test (hCXCR4 vs LacZ)
0	2.33	0.028
0.5	1.10	0.344
10	0.94	0.208

**Legend** "Fold Change" indicates the following ratio; (Mean % of hCXCR4-expressing cells incubated with 20 nM SDF1 for 24 hours that were scored positive for cell surface Glut4)/(Mean % of LacZ-expressing cells cells incubated with 20 nM SDF1 for 24 hours that were scored positive for cell surface Glut4). "h" is human. "n" is the number of experiments.

[357] The results show that in a cell such as an adipocyte, increasing the levels of CXCR4 in the presence of the CXCR4 ligand SDF1 leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of CXCR4 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[358] CXCR4 contains the following protein domains (designated with reference to SEQ ID NO:70): 7 transmembrane receptor (Secretin family) (PF00002) at amino acids 45 to 280; and 7 transmembrane domains (TMHMM2.0) at amino acids 43 to 65, 78 to 96, 111 to 132, 155 to 174, 199 to 221, 242 to 264, 284 to 306. CXCR4 is a G protein-coupled receptor that binds the CXC cytokine, CXCL12. CXCR4 may be required for hematopoiesis and organ vascularization (Tachibana, K. *et al.*, *Nature* 393: 591-4(1998). It is known to act as a coreceptor for HIV (Moriuchi, M.*et al.*, *J. Immunol.* 159: 4322-4329 (1997)) and inhibition of this receptor may be therapeutic for invasive breast cancer (Tamamura, H., *et al.*, *FEBS Lett.* 550: 79-83 (2003)).

[359] Stimulation of the CXCR4 receptor with its ligand SDF1-alpha leads to an increase in intracellular Ca<sup>2+</sup> (see, e.g., Princen K. *et al. J Exp Med.* 20;186: 1383-1388 (1997)). Agonists of the CXCR4 receptor can therefore be identified, e.g., by screening cells with high levels of the CXCR4 receptor to identify compounds that increases intracellular Ca<sup>2+</sup> using a calcium sensitive dye such as Calcium 3 or Fluo 3.

**DDAH2**

[360] Probe set 214909 detects DDAH2 nucleic acid sequences. Expression of DDAH2 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	127.86	3.27	10	96.19	6.67	8	1.33	0.002	DDAH2

- 5 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- 10 [361] Probe set 214909 detects DDAH2 nucleic acid sequences. Expression of DDAH2 transcripts was increased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

B/C	CORRELATION RD			Students t test	Gene Name
	n	Corr Co-efficient			
C	26	-0.814		<0.005	DDAH2

B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

- 15 [362] DDAH2 was also evaluated using real-time PCR. The results further show that DDAH2 is significantly increased in patients with insulin resistance compared to normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (26)	-0.556	<0.005

- 20 "Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[363] Probe set 214909 detects DDAH2 nucleic acid sequences. Expression of DDAH2 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
100.24	8.97	5	29.85	4.51	13	3.36	<0.000	DDAH2

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [364] DDAH2 contains the following protein domains (designated with reference to SEQ ID NO:76): Amidinotransferase (PF02274) at amino acids 6 to 281. DDAH2 regulates cellular methylarginine concentrations, which in turn inhibit nitric oxide synthase. DDAH2 expression predominates in more highly vascularized tissues and in immune tissues (Leiper, J.M., *et al.*, *Biochem J.* 343: 209-14 (1999)).

### DERP7

- 10 [365] Probe set 219410 detects DERP7 nucleic acid sequences. Expression of DERP7 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN					
	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
B	53.96	6.71	5	24.77	0.29	4	2.18	0.012	DERP7

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

- 15 Change" indicates fold change of obese in comparison to lean patients.

[366] DERP7 was also evaluated using real-time PCR. The results further show that DERP7 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	3.18	0.074

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese

- 20 expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[367] Probe set 219410 detects DERP7 nucleic acid sequences. Expression of DERP7 transcripts was decreased in pio compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	PIO			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	278.73	15	3	432.9	27.78	3	0.64	0.015	DERP7

B/C indicates sample is from Basal or Clamp; "Pre-Pio" and "Post-Pio" indicates sample was taken before or after 24 hours of pioglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-pio in comparison to pre-pio samples.

[368] DERP7 contains the following protein domains (designated with reference to SEQ ID NO:82): Family of unknown function (DUF716) (PF04819) at amino acids 113 to 251; and 7 transmembrane domains (TMHMM2.0) at amino acids 4 to 23, 44 to 66, 91 to 113, 118 to 140, 150 to 172, 181 to 203, 218 to 240. DERP7 has high similarity to an uncharacterized mouse protein, p.19.5. This is a putative membrane protein which was shown to be differentially expressed in two closely related T lymphoma cell clones (MacLeod C.L. *et al.*, *Cell Growth Differ.*, 1(6): 271-279 (1990)).

### ENDOGLYX1

[369] Probe set 219091 detects ENDOGLYX1 nucleic acid sequences.

Expression of ENDOGLYX1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	221.54	17.86	5	146.55	13.61	4	1.51	0.013	ENDOGLYX1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[370] ENDOGLYX1 was also evaluated using real-time PCR. The results further show that ENDOGLYX1 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	2.36	0.018

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese-expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[371] Probe set 219091 detects ENDOGLYX1 nucleic acid sequences.

Expression of ENDOGLYX1 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
151.52	16.51	5	40.09	5.18	13	3.78	0.002	ENDOGLYX1

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

- 5 indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [372] ENDOGLYX1 contains the following protein domains (designated with reference to SEQ ID NO:88): Seryl-tRNA synthetase N-terminal domain (PF02403) at amino acids 499 to 600; Myosin tail (PF01576) at amino acids 143 to 814; Apolipoprotein A1/A4/E family (PF01442) at amino acids 200 to 469; C1q domain (PF00386) at amino acids 827 to 946; TNF(Tumour Necrosis Factor) family (PF00229) at amino acids 835 to 946; and Intermediate filament protein (PF00038) at amino acids 360 to 645. ENDOGLYX1 is a cell surface glycoprotein which is attached to the extracellular matrix and capable of forming homo- and heteromers via disulfide bonding. It may play a role in angiogenesis, vasculogenesis, cell-matrix adhesion, and hemostasis (Christian, S. *et al.*, *J Biol Chem* 276: 48588-95 (2001); Leimeister, C., *et al.*, *Dev Biol.* 249: 204-18 (2002)).

### ETL

[373] Probe set MBXHUMFAT01286 detects ETL nucleic acid sequences.

- 20 Expression of ETL transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN					
	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
C	48.6	4.32	8	32.99	3.67	8	1.47	0.016	ETL

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[374] Probe set MBXHUMFAT01286 detects ETL nucleic acid sequences.

- 25 Expression of ETL transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
17.82	5.34	5	5.36	1.22	13	3.32	0.079	ETL

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [375] ETL was also evaluated using real-time PCR. The results further show that ETL is significantly over-expressed in adipose tissues when compared to all other human adult tissues.

Comparison	Expression Fold Change	t test
Fat Tissues (5)/ All Other Tissues (13)	5.15	0.007

"Fold Change" indicates the fold expression calculated as the ratio of the mean adipose tissues expression/ mean other tissues expression. Numbers in parentheses indicates the number of human adult tissue samples analyzed by real-time PCR.

- [376] ETL contains the following protein domains (designated with reference to SEQ ID NO:94): BphX-like (PF06139) at amino acids 629 to 728; Latrophilin/CL-1-like GPS domain (PF01825) at amino acids 487 to 539; EGF-like domain (PF00008) at amino acids 183 to 220; 7 transmembrane receptor (Secretin family) (PF00002) at amino acids 545 to 792; and 7 transmembrane domains (TMHMM2.0) at amino acids 552 to 574, 587 to 606, 621 to 643, 655 to 677, 692 to 714, 740 to 762, 766 to 788. ETL belongs to the secretin family of G-protein-coupled peptide hormone receptors and the EGF-TM7 subfamily of receptors. The latter are characterized by a variable number of extracellular EGF and cell surface domains and conserved seven transmembrane-spanning regions. (Nechiporuk, T. *et al.*, *J Biol Chem* 276: 4150-7 (2001)).

## 20 FLJ12389

[377] Probe set 218434 detects FLJ12389 nucleic acid sequences. Expression of FLJ12389 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN					
	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
C	88.31	9.03	10	329.7	53.82	8	0.27	0.003	FLJ12389

- B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.



[378] FLJ12389 was also evaluated using real-time PCR. The results further show that FLJ12389 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.19	0.001

- “Fold Change” indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[379] Probe set 218434 detects FLJ12389 nucleic acid sequences. Expression of FLJ12389 transcripts was decreased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	182.13	24.09	8	329.7	53.82	8	0.55	0.032	FLJ12389

- 10 B/C indicates sample is from Basal or Clamp; “Mean Expr” indicates mean expression; “SEM” indicates standard error of mean; “n” indicates number of patient samples; “Fold Change” indicates fold change of obese in comparison to lean patients.

- [380] FLJ12389 was also evaluated using real-time PCR. The results further show that FLJ12389 is significantly under-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (8) / Lean (8)	0.49	0.014

“Fold Change” indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- 20 [381] Probe set 218434 detects FLJ12389 nucleic acid sequences. Expression of FLJ12389 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES			Fold Change	Students t test	Gene Name
Mean Expr	SEM	n	Mean Expr	SEM	n			
266.24	76.69	5	48.32	9.58	13	5.51	0.046	FLJ12389

“Mean Expr” indicates mean expression; “SEM” indicates standard error of mean; “n” indicates number of tissue samples; “Fold Change” indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

[382] FLJ12389 contains the following protein domains (designated with reference to SEQ ID NO:100): AMP-binding enzyme (PF00501) at amino acids 130 to 571. FLJ12389 has some sequence similarity to acetyl coenzyme A synthetases and is predicted to contain ATP/GTP and AMP binding sites. FLJ12389 may be a ketone body-utilizing enzyme of which the physiological role of remains unclear (Ohgami, M. *et al.*, *Biochem Pharmacol* 65: 989-994 (2003)).

#### **FZD4**

[383] Probe set 218665 detects FZD4 nucleic acid sequences. Expression of FZD4 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	425.71	13.15	10	543.05	29.78	8	0.78	0.005	FZD4

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[384] FZD4 was also evaluated using real-time PCR. The results further show that FZD4 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.82	0.016

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[385] Probe set 218665 detects FZD4 nucleic acid sequences. Expression of FZD4 transcripts was decreased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

B/C	CORRELATION RD			Gene Name
	n	Corr Co-efficient	Students t test	
C	18	0.655	<0.005	FZD4

B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

[386] FZD4 was also evaluated using real-time PCR. The results further show that FZD4 is significantly decreased in patients with insulin resistance compared to normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (18)	0.651	<0.005

"Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[387] Probe set 218665 detects FZD4 nucleic acid sequences. Expression of FZD4 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
312.52	56.2	5	65.81	6.04	13	4.75	0.011	FZD4

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

[388] FZD4 contains the following protein domains (designated with reference to SEQ ID NO:110): Signal peptide at amino acids 1 to 36; Frizzled/Smoothed family membrane region (PF01534) at amino acids 209 to 514; Fz domain (PF01392) at amino acids 35 to 159; and 7 transmembrane domains (TMHMM2.0) at amino acids 10 to 32, 221 to 243, 253 to 275, 301 to 323, 394 to 416, 437 to 459, 474 to 496. FZD4 encodes a 7-transmembrane domain protein and is a receptor for Wnt signaling proteins. The auditory and cerebellar phenotypes of FZD4 null mice implicate Frizzled signaling in maintaining the viability and integrity of the nervous system in later life and retinal angiogenesis (Wang, Y., *et al.*, *J Neurosci.* 21: 4761-71 (2001); Singaraja, R.R., *et al.*, *Nat Genet.* 32: 326-30 (2002)).

### GLIPR1

[389] Probe set 226142 detects GLIPR1 nucleic acid sequences. Expression of GLIPR1 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	45.5	4.89	10	18.31	2.44	8	2.48	<0.000	GLIPR1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [390] GLIPR1 was also evaluated using real-time PCR. The results further show that GLIPR1 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	2	0.017

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [391] Probe set 226142 detects GLIPR1 nucleic acid sequences. Expression of GLIPR1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	33.18	5.13	5	14.48	2.07	4	2.29	0.018	GLIPR1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[392] GLIPR1 was also evaluated using real-time PCR. The results further show that GLIPR1 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5)/ Lean (4)	2.93	0.015

- "Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[393] Probe set 226142 detects GLIPR1 nucleic acid sequences. Expression of GLIPR1 transcripts was decreased in pio+insulin compared to insulin treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	PIO+INSULIN			INSULIN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	73.17	5.25	3	114.93	4.09	3	0.64	0.004	GLIPR1

B/C indicates sample is from Basal or Clamp; "Pre-Pio" and "Post-Pio" indicates sample was taken before or after 24 hours of pioglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-pio in comparison to pre-pio samples.

[394] Probe set 226142 detects GLIPR1 nucleic acid sequences. Expression of GLIPR1 transcripts was decreased in rosi+insulin compared to insulin treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	ROSI+INSULIN			INSULIN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	69.57	3.72	3	114.93	4.09	3	0.61	0.001	GLIPR1

B/C indicates sample is from Basal or Clamp; "Pre-Rosi" and "Post-Rosi" indicates sample was taken before or after 24 hours of rosiglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-rosi in comparison to pre-rosi samples.

[395] GLIPR1 contains the following protein domains (designated with reference to SEQ ID NO:118): Signal peptide at amino acids 1 to 21; SCP-like extracellular protein (PF00188) at amino acids 38 to 174; and 1 transmembrane domain (TMHMM2.0) at amino acids 235 to 257. GLIPR1 is a putative secreted protein that may play a role in inhibition of malignant growth and progression through its proapoptotic activities (Ren, C. *et al.*, *Mol Cell Biol* 22: 3345-57 (2002)).

## 20 GPR105

[396] Probe set 206637 detects GPR105 nucleic acid sequences. Expression of GPR105 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	98.55	8.61	10	60.97	3.98	8	1.62	0.002	GPR105

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [397] GPR105 was also evaluated using real-time PCR. The results further show that GPR105 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	2	0.007

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [398] GPR105 contains the following protein domains (designated with reference to SEQ ID NO:124): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 39 to 295; and 7 transmembrane domains (TMHMM2.0) at amino acids 27 to 49, 56 to 78, 98 to 117, 137 to 159, 187 to 209, 235 to 257, 279 to 298. GPR105 is a G(i/o)- G protein-coupled receptor that is activated by extracellular UDP-sugars. Activation of the receptor stimulates intracellular calcium and may mediate primitive hematopoietic cell responses to microenvironments (Chambers, J. K. *et al.*, *J Biol Chem* 275: 10767-71 (2000); Lee, B. C. *et al.*, *Genes Dev* 17: 1592-604 (2003); Skelton, L. *et al.*, *J Immunol* 171: 1941-9 (2003); Moore, D. J. *et al.*, *Brain Res Mol Brain Res* 118: 10-23 (2003)).

#### GPR146

- [399] Probe set 228770 detects GPR146 nucleic acid sequences. Expression of GPR146 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	42.73	2.15	10	57.98	4.47	8	0.74	0.012	GPR146

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[400] GPR146 was also evaluated using real-time PCR. The results further show that GPR146 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.66	0.014

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[401] Probe set 228770 detects GPR146 nucleic acid sequences. Expression of GPR146 transcripts was decreased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

CORRELATION RD				
B/C	n	Corr Co-efficient	Students t test	Gene Name
C	26	0.679	<0.005	GPR146

B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

[402] GPR146 was also evaluated using real-time PCR. The results further show that GPR146 is significantly decreased in patients with insulin resistance compared to normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (26)	0.535	<0.005

"Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[403] The cellular level of GPR146 was reduced in 3T3-L1 adipocytes using siRNA directed against GPR146 and the effect on basal and insulin stimulated glucose transport was determined

**GPR146 mRNA Level in 3T3-L1 Adipocytes Transfected with siRNA Oligonucleotides**

	Expt 1	Expt 2	Expt 3	Expt 4	Mean FC ± SEM (n=4)
FC (siRNA/Scr)	0.545	0.443	0.399	0.524	0.478 ± 0.03

**Legend:** "siRNA" indicates Dharmacon Smartpool siRNA oligonucleotides directed against murine GPR146. "Scr" indicates the Dharmacon Scramble siRNA Control. "FC" indicates the fold change defined as the following ratio; Level of GPR146 mRNA in GPR146 siRNA transfected 3T3-L1 adipocytes/Level of GPR146 mRNA in Scramble siRNA transfected 3T3-L1 adipocytes. "n" is the number of experiments. SEM is the standard error of the mean.

**Glucose Transport in 3T3-L1 Adipocytes Transfected with siRNA Oligonucleotides**

	Experiment 1		Experiment 2		Experiment 3		Experiment 4		
Insulin (nM)	FC (siRNA A/Scr)	t-test	FC (siRNA A/Scr)	t-test	FC (siRNA A/Scr)	t-test	FC (siRNA A/Scr)	t-test	Mean FC (siRNA/S ± SEM (n=4)
0.0	0.637	<0.001	0.614	<0.001	0.688	<0.001	0.785	0.0461	0.681 ± 0.03
0.1	0.488	<0.001	0.479	<0.001	0.635	<0.001	0.679	<0.001	0.570 ± 0.03
1.0	0.496	<0.001	0.609	<0.001	0.845	<0.001	0.886	0.0314	0.709 ± 0.03
10.0	0.546	<0.001	0.615	<0.001	0.803	<0.001	0.908	0.0669	0.718 ± 0.03

**Legend:** "siRNA" indicates Dharmacon Smartpool siRNA oligonucleotides directed against murine GPR146. "Scr" indicates Dharmacon Scramble siRNA Control oligonucleotides. "FC" indicates the fold change defined as the following ratio; glucose transport in GPR146 siRNA transfected cells/glucose transport in Scramble siRNA transfected cells. "n" is the number of experiments. SEM is the standard error of the mean.

[404] The results show that decreasing the levels of GPR146 in a cell such as an adipocyte leads to a corresponding decrease in glucose uptake. This indicates that increasing the levels of GPR146 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[405] GPR146 contains the following protein domains (designated with reference to SEQ ID NO:130): Signal peptide at amino acids 1 to 37; and a 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 44 to 296. GPR146 is a member of the rhodopsin family of G protein-coupled receptors (GPCR) with very low sequence similarity to GPR30.



- [406] GPR146 is an orphan GPCR with unknown coupling to G proteins (Goriam DE et al. *Biochim Biophys Acta*. 1722:235-46 (2005)). Agonists of GPR146 can be detected, e.g., using cells over-expressing GPR146 along with either a promiscuous G protein (such as Gα16, Kostenis E. *Trends Pharmacol Sci*. 22:560-4 (2001) or a chimeric G protein (such as Gα16z or Gα16s) (see, e.g., Liu AM et al. *J Biomol Screen*. 8:39-49 (2003), Hazari A et al. *Cell Signal*. 16:51-62 (2004)). Screening of such cells will detect GPR146 agonists, for example either by measuring intracellular Ca<sup>2+</sup> with a Ca<sup>2+</sup> sensitive dye such as Calcium 3 or by measuring intracellular cyclic AMP.

### GPR30

- 10 [407] Probe set 210640 detects GPR30 nucleic acid sequences. Expression of GPR30 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	57.91	2.58	10	68.68	1.19	8	0.84	0.002	GPR30

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

- 15 Change" indicates fold change of diabetics in comparison to lean patients.

[408] GPR30 was also evaluated using real-time PCR. The results further show that GPR30 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.79	0.005

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic

- 20 expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[409] Probe set 210640 detects GPR30 nucleic acid sequences. Expression of GPR30 transcripts was decreased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

CORRELATION RD				
B/C	n	Corr Co-efficient	Students t test	Gene Name
C	26	0.608	<0.005	GPR30

- 25 B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates

down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

[410] GPR30 was also evaluated using real-time PCR. The results further show that GPR30 is significantly decreased in patients with insulin resistance compared to normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (26)	0.5	<0.01

"Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[411] Probe set 210640 detects GPR30 nucleic acid sequences. Expression of GPR30 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
48.76	7.76	5	13.04	4.3	13	3.74	0.006	GPR30

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

[412] GPR30 contains the following protein domains (designated with reference to SEQ ID NO:136): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 76 to 324; and 7 transmembrane domains (TMHMM2.0) at amino acids 63 to 85, 97 to 119, 134 to 153, 174 to 196, 219 to 241, 262 to 284, 304 to 326. GPR30 a G protein-coupled receptor that stimulates adenylyl cyclase and mediates attenuation of Erk-1/-2 activity by estrogen via Raf-1 inactivation. GPR30 is a progestin target gene whose expression correlates with progestin-induced growth inhibition in breast cancer cells (Filardo, E. J. *et al.*, *Mol Endocrinol* 14: 1649-60 (2000); Ahola, T. M. *et al.*, *Endocrinology* 143: 4620-6 (2002)).

## 25 GPR65

[413] Probe set 214467 detects GPR65 nucleic acid sequences. Expression of GPR65 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	11.44	0.48	5	7.23	0.95	4	1.58	0.013	GPR65

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [414] GPR65 was also evaluated using real-time PCR. The results further show that GPR65 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	4.79	0.012

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [415] GPR65 contains the following protein domains (designated with reference to SEQ ID NO:142): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 31 to 290; and 7 transmembrane domains (TMHMM2.0) at amino acids 15 to 37, 49 to 71, 91 to 110, 130 to 152, 181 to 203, 224 to 246, 271 to 293. GPR65 is a G protein-coupled receptor activated by the glycosphingolipid psychosine. It may function in apoptosis and immunological autotolerance, and plays a role in T-cell associated diseases and Globoid cell leukodystrophy (Kyaw, H. *et al.*, *DNA Cell Biol* 17: 493-500 (1998); Im, D. S. *et al.*, *J Cell Biol* 153: 429-34 (2001)). GPR65 has been described to be expressed in human primary monocytes and macrophages (Duong, C.Q., *et al.*, *Biochim Biophys Acta* 1682: 112-9(2004)).

## HTR2B

- [416] Probe set 206638 detects HTR2B nucleic acid sequences. Expression of HTR2B transcripts was decreased in pio compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	PIO			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	6.13	1.65	3	15.9	1.01	3	0.39	0.012	HTR2B

B/C indicates sample is from Basal or Clamp; "Pre-Pio" and "Post-Pio" indicates sample was taken before or after 24 hours of pioglitazone treatment; "Mean Expr" indicates mean

- expression; "SEM" indicates standard error of mean; "n" indicates number of patient

samples; "Fold Change" indicates fold change of primary human adipocytes post-pio in comparison to pre-pio samples.

[417] HTR2B was also evaluated using real-time PCR. The results further show that HTR2B is significantly under-expressed in primary cultured human adipocytes treated with pio when compared to vehicle.

Comparison	Expression Fold Change	t test
Post-Pio (12) / Pre-Pio (12)	0.63	0.002

"Fold Change" indicates the fold expression calculated as the ratio of the mean pio expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

[418] HTR2B was over-expressed in 3T3-L1 adipocytes and the cells were then treated with 5-HT and the effect on basal and insulin stimulated glucose transport was determined.

Glucose Transport Analysis in 3T3-L1 Adipocytes:

	Experiment 1		Experiment 2		Experiment 3		
Insulin (nM)	FC (hHTR2B/Con)	t test ( $\alpha$ )	FC (hHTR2B/Con)	t test ( $\alpha$ )	FC (hHTR2B/Con)	t test ( $\alpha$ )	Mean FC (hHTR2B/Con) +/- SEM (n=3)
0	1.95	0.01	1.51	0.001	1.38	0.02	1.61±0.17
0.03	1.65	0.03	1.15	0.39	1.47	0.01	1.42±0.15
0.3	1.26	0.04	1.12	0.01	1.30	0.10	1.22±0.05
3	1.10	0.17	1.09	0.09	1.02	0.77	1.07±0.03

**Legend:** "Con" indicates control 3T3-L1 adipocytes that do not express hHTR2B. "FC" indicates the fold change defined as the following ratio; glucose transport in HTR2B-expressing cells stimulated with 1  $\mu$ M 5HT for 3 hours/glucose transport in non-HTR2B-expressing cells stimulated with 1  $\mu$ M 5HT for 3 hours. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

[419] These results show that in a cell such as a adipocyte, increasing the levels of HTR2B in the presence of its ligand 5-HT leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of HTR2B in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[420] HTR2B contains the following protein domains (designated with reference to SEQ ID NO:148): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 71 to 380; and 7 transmembrane domains (TMHMM2.0) at amino acids 59 to 81, 93 to 115, 130 to 149, 170 to 192, 217 to 239, 327 to 349, 364 to 383. HTR2B is a 5-

hydroxytryptamine 2B (serotonin) receptor that signals through phospholipase C and is known to induce mitogenesis, mediate contractile effects of serotonin on GI tract smooth muscle and may play a role in digestion and migraine headaches (Duxon M.S. *et al.*, *Neuroscience* 76(2): 323-9 (1997); Jerman J.C. *et al.*, *Eur J Pharmacol.*, 414(1): 23-30 (2001); Launay, J.M., *et al.*, *J Biol Chem.* 271: 3141-7 (1996); Nebigil, C.G. *et al.*, *Proc Natl Acad Sci U S A.* 97: 2591-6 (2000); Schaerlinger, B., *et al.*, *Br J Pharmacol.* 140: 277-84. Epub (2003)).

- [421] The HTR2B receptor is a G protein coupled receptor linked to the mobilization of intracellular  $Ca^{2+}$  (*see, e.g.*, Schmuck K. *et al FEBS Lett.* 342 :85-90 (1994)).
- 10 Agonists of the HTR2B receptor can therefore be identified, *e.g.*, by screening cells with high levels of the HTR2B receptor to identify compounds that increase intracellular  $Ca^{2+}$  using a calcium sensitive dye such as Calcium 3 or Fluo 3.

### ITGB2

- [422] Probe set 202803 detects ITGB2 nucleic acid sequences. Expression of
- 15 ITGB2 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

DIABETIC				LEAN					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
C	143.11	20.64	10	63.33	10.52	8	2.26	0.004	ITGB2

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- 20 [423] ITGB2 was also evaluated using real-time PCR. The results further show that ITGB2 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	2.41	0.02

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [424] Probe set 202803 detects ITGB2 nucleic acid sequences. Expression of ITGB2 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	133.4	23.81	5	49.3	8.55	4	2.71	0.021	ITGB2

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [425] ITGB2 was also evaluated using real-time PCR. The results further show that ITGB2 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5)/ Lean (4)	5.65	0.05

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

10

[426] ICAM-1, the ligand for ITGB2-containing integrin complexes such as LFA-1, was added to cultures of 3T3-L1 adipocytes and the effect on glucose transport and Glut 4 translocation was determined.

#### Glut4 Translocation Analysis:

Insulin (nM)	Fold Change (ICAM + Mn2+/ Mn2+) (n=3)	t test (ICAM + Mn vs Mn)
0	0.55	0.002
0.5	0.79	0.006
10	0.95	0.221

- 15 Legend "Fold Change" indicates the following ratio; (Mean % of 3T3-L1 adipocytes incubated 3 hours with 10 ug/ml ICAM + 200 uM Mn2+ scored positive for cell surface Glut4)/(Mean % 3T3-L1 adipocytes incubated 3 hours with 200 uM Mn2+ scored positive for cell surface Glut4). "h" is human. "n" is the number of experiments.

- 20 [427] Incubating 3T3-L1 adipocytes with 200 uM Mn2+ for 3 hours enhanced cell surface Glut4 in the absence of insulin. In contrast, including 10 ug/ml ICAM inhibited the increase observed with Mn alone.

Glucose Transport Analysis in 3T3-L1 Adipocytes:

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (ICAM+Mn/Mn) +/- SD (n=3)
	FC (ICAM+ Mn / Mn)	t test ( $\alpha$ )	FC (ICAM+ Mn / Mn)	t test ( $\alpha$ )	FC (ICAM+ Mn / Mn)	t test ( $\alpha$ )	
0	0.71	0.39	0.66	0.15	0.73	0.006	0.70±0.04
0.03	0.63	0.03	0.77	0.13	0.73	0.003	0.71±0.07
0.3	0.69	0.34	0.81	0.05	0.69	0.002	0.73±0.07
3	0.76	0.28	0.72	0.24	0.67	0.008	0.72±0.05

**Legend:** "FC" indicates the fold change defined as the following ratio; glucose transport in 3T3-L1 adipocytes incubated with 10 ug/ml ICAM1 + 200 uM Mn for 3 hours/glucose transport in 3T3-L1 adipocytes incubated with 200 uM Mn for 3 hours. "h" is human. "n" is the number of experiments. SD is the standard deviation.

[428] The results show that increasing the activity of ITGB2 containing integrins such as LFA-1 in a cell such as an adipocyte leads to a corresponding decrease in glucose uptake. This indicates that decreasing the levels or activity of ITGB2-containing integrins such as LFA-1 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[429] ITGB2 contains the following protein domains (designated with reference to SEQ ID NO:154): Signal peptide at amino acids 1 to 22; Plexin repeat (PF01437) at amino acids 24 to 63; Integrin, beta chain (PF00362) at amino acids 32 to 447; EGF-like domain (PF00008) at amino acids 582 to 612; and 1 transmembrane domain (TMHMM2.0) at amino acids 701 to 723. The ITGB2 protein product is the integrin beta chain beta 2. Integrins are integral cell-surface proteins composed of an alpha chain and a beta chain. A given chain may combine with multiple partners resulting in different integrins. For example, beta 2 combines with the alpha L chain to form the integrin LFA-1, and combines with the alpha M chain to form the integrin Mac-1. Integrins are known to participate in cell adhesion as well as cell-surface mediated signalling. The regulation of interaction mediated by adhesion molecules may provide new targets for controlling inflammatory and immune responses (Bunting, M. *et al.*, *Curr Opin Hematol.* 9: 30-5 (2002); Tsuji T. *et al.*, *Blood* 91(4): 1263-71 (1998); Zhang L. and Plow E.F. *Biochemistry* 38(25): 8064-71 (1999)).

[430] Inhibitors of LFA-1 can be detected using a variety of assays, such as those that detect the ability of candidate compounds to disrupt the association of LFA-1 with

the ligand ICAM-1. For example, purified LFA-1 can be immobilized and compounds that cause the inhibition of biotinylated ICAM-1 binding can be detected.

[431] Inhibitors of the LFA-1 integrin complex (alphaL/beta2) have been developed. These include inhibitors such as BIRT0377, LFA703 and A-286982 (Shimakoa *et al Nat. Rev. Drug Discovery* 2: 703-716 (2003). Such inhibitors, as well as other inhibitors of LFA-1 integrin complex, are useful as agents for the treatment for insulin resistance and diabetes.

### ITIH5

[432] Probe set MBXHUMFAT04252 detects ITIH5 nucleic acid sequences.

- 10 Expression of ITIH5 transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	60.99	4.04	10	30.15	6.12	8	2.02	0.001	ITIH5

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- 15 [433] ITIH5 was also evaluated using real-time PCR. The results further show that ITIH5 is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.78	0.001

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient

- 20 samples analyzed by real-time PCR.

[434] Probe set MBXHUMFAT04252 detects ITIH5 nucleic acid sequences.

Expression of ITIH5 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	48.05	5.62	8	30.15	6.12	8	1.59	0.049	ITIH5

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

- 25 "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.



[435] ITIH5 was also evaluated using real-time PCR. The results further show that ITIH5 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (8)/ Lean (8)	1.43	0.094

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese

- 5 expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[436] Probe set MBXHUMFAT04252 detects ITIH5 nucleic acid sequences. Expression of ITIH5 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES			Fold Change	Students t test	Gene Name
Mean Expr	SEM	n	Mean Expr	SEM	n			
19.18	4.15	5	1.86	0.63	13	10.3	0.013	ITIH5

10

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled

- [437] ITIH5 was over-expressed in 3T3-L1 adipocytes and the effect on  
 15 basal and insulin stimulated glucose transport and Glut 4 translocation was determined  
Glucose Transport Analysis in 3T3-L1 Adipocytes:

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (hITIH5/Con) +/- SEM (n=3)
	FC (hITIH5 / Con)	t test ( $\alpha$ )	FC (hITIH5 / Con)	t test ( $\alpha$ )	FC (hITIH5 / Con)	t test ( $\alpha$ )	
0	1.29	0.410	1.46	0.065	1.32	0.227	1.36±0.05
0.05	1.24	0.316	1.36	0.058	1.67	0.009	1.42±0.13
0.1	1.11	0.054	1.29	0.027	1.18	0.246	1.19±0.05
0.3	0.87	0.104	1.18	0.109	1.18	0.063	1.08±0.1
1	1.13	0.293	1.11	0.084	1.13	0.224	1.12±0.01
10	1.14	0.327	1.03	0.498	0.94	0.153	1.04±0.06

Legend: "Con" indicates control 3T3-L1 adipocytes that do not express ITIH5. "FC" indicates the fold change defined as the following ratio; glucose transport in ITIH5-expressing 3T3-L1 adipocytes /glucose transport in non-PTPRE-expressing 3T3-L1 adipocytes. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

20

Glut4 Translocation Analysis:

Insulin (nM)	Fold Change (Mean hITI5/Mean LacZ) (n=3)	t test (hITI5 vs LacZ)
0	2.43	0.004
0.5	1.0	0.422
10	0.96	0.170

Legend "Fold Change" indicates the following ratio; (Mean % of ITIH5-expressing cells scored positive for cell surface Glut4)/(Mean % of LacZ-expressing cells scored positive for cell surface Glut4). "h" is human. "n" is the number of experiments.

5 [438] The results show that increasing the levels of ITIH5 in a cell such as an adipocyte leads to a corresponding increase in glucose uptake. This indicates that increasing the levels or activity of ITIH5 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective  
10 treatment for insulin resistance and diabetes.

[439] ITIH5 contains the following protein domains (designated with reference to SEQ ID NO:160): Signal peptide at amino acids 1 to 21; Inter-alpha-trypsin inhibitor heavy chain C-terminus (PF06668) at amino acids 715 to 909; T-box (PF00907) at amino acids 310 to 426; and von Willebrand factor type A domain (PF00092) at amino acids  
15 295 to 478. A soluble active secreted form of ITIH5 has been detected and this is displayed in SEQ ID NO:161. ITIH5 belongs to the inter-alpha-trypsin inhibitor (ITI) family constitutes a group of proteins built up from one light chain and a variable set of heavy chains. Originally identified as plasma protease inhibitors, recent data indicate that ITI proteins play a role in extracellular matrix (ECM) stabilization and in prevention of tumor metastasis. ITIH5  
20 expression was found to be consistently downregulated in invasive mammary ductal carcinomas (Himmelfarb M. *et al.*, *Cancer Lett.* 204(1): 69-77 (2004)).

LGALS12

[440] Probe set 223828 detects LGALS12 nucleic acid sequences.

Expression of LGALS12 transcripts was decreased in diabetic compared to lean patients in  
25 the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	106.03	11	10	283.66	39.14	8	0.37	0.002	LGALS12

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[441] LGALS12 was also evaluated using real-time PCR. The results further show that LGALS12 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.31	<0.000

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic

- 5 expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[442] Probe set 223828 detects LGALS12 nucleic acid sequences.

Expression of LGALS12 transcripts was decreased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	153.48	16.5	8	283.66	39.14	8	0.54	0.013	LGALS12

- 10 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [443] LGALS12 was also evaluated using real-time PCR. The results further show that LGALS12 is significantly under-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (8) / Lean (8)	0.49	0.002

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[444] Probe set 223828 detects LGALS12 nucleic acid sequences.

- 20 Expression of LGALS12 transcripts was decreased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	108.02	16.97	5	163.5	9.22	4	0.66	0.028	LGALS12

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[445] Probe set 223828 detects LGALS12 nucleic acid sequences.

Expression of LGALS12 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
150.4	40.21	5	7.31	2.04	13	20.58	0.024	LGALS12

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

- 5 indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

[446] LGALS12 contains the following protein domains (designated with reference to SEQ ID NO:165): Galactoside-binding lectin (PF00337) at amino acids 48 to 182. LGALS12 is a member of the beta-galactoside-binding lectin family and may be an apoptosis activator that negatively regulates the cell cycle and cell proliferation (Yang R.Y. *et al.*, *J Biol Chem.*, 276(23): 20252-60 (2001); Hotta K. *et al.*, *J Biol Chem.*, 276(36): 34089-97 (2001); Liu, F.T., *et al.*, *Biochim Biophys Acta.* 1572: 263-73 (2002)).

#### NMB

- [447] Probe set 205204 detects NMB nucleic acid sequences. Expression of NMB transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN					
	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
C	173.2	11.31	10	144.88	10.76	8	1.2	0.089	NMB

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

Change" indicates fold change of diabetics in comparison to lean patients.

- 20 [448] Probe set 205204 detects NMB nucleic acid sequences. Expression of NMB transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
188.38	32.19	5	35.72	2.98	13	5.27	0.009	NMB

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [449] NMB was also evaluated using real-time PCR. The results further show that NMB is significantly over-expressed in adipose tissues when compared to all other human adult tissues.

Comparison	Expression Fold Change	t test
Fat Tissues (5)/ All Other Tissues (13)	26.71	0.029

"Fold Change" indicates the fold expression calculated as the ratio of the mean adipose tissues expression/ mean other tissues expression. Numbers in parentheses indicates the number of human adult tissue samples analyzed by real-time PCR.

- [450] NMB contains the following protein domains (designated with reference to SEQ ID NO:181): Signal peptide at amino acids 1 to 26; and Bombesin-like peptide (PF02044) at amino acids 47 to 60. A soluble active secreted form of NMB has been detected and this is displayed in SEQ ID NO:182. NMB is a bombesin-related neuropeptide which induces calcium flux and phosphatidylinositol turnover leading to stimulation of cell proliferation through the G-protein coupled neuromedin B receptor (Ohki-Hamazaki H. *Prog. Neurobiol.* 62(3): 297-312 (2000); Mason S. *et al.*, *Eur J Pharmacol.*, 438(1-2): 25-34 (2002)).

### NNAT

- [451] Probe set 204239 detects NNAT nucleic acid sequences. Expression of NNAT transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN					
	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
C	80.57	7.75	10	122.46	11.93	8	0.66	0.012	NNAT

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

[452] NNAT was also evaluated using real-time PCR. The results further show that NNAT is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.64	0.14

5 "Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[453] Probe set 204239 detects NNAT nucleic acid sequences. Expression of NNAT transcripts was increased in rosi compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	ROSI			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	36.43	2.23	3	18.67	2.29	3	1.95	0.005	NNAT

10 B/C indicates sample is from Basal or Clamp; "Pre-Rosi" and "Post-Rosi" indicates sample was taken before or after 24 hours of rosiglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-rosi in comparison to pre-rosi samples.

15 [454] NNAT was also evaluated using real-time PCR. The results further show that NNAT is significantly over-expressed in primary cultured human adipocytes treated with rosi when compared to vehicle.

Comparison	Expression Fold Change	t test
Post-Rosi (12)/ Pre-Rosi (12)	2.13	0.004

20 "Fold Change" indicates the fold expression calculated as the ratio of the mean rosi expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

[455] Probe set 204239 detects NNAT nucleic acid sequences. Expression of NNAT transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
125	14.04	5	32.15	7.89	13	3.89	0.001	NNAT

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [456] NNAT contains the following protein domains (designated with reference to SEQ ID NO:188): Signal peptide at amino acids 1 to 23; and 1 transmembrane domain (TMHMM2.0) at amino acids 13 to 35. NNAT is a putative proteolipid that may regulate ion channels during brain development. It is found to be highly expressed in pituitary adenomas and is frequently hypermethylated in childhood myeloid and lymphoid acute leukemias (Dou D. and Joseph R. *Genomics* 33(2): 292-7 (1996); Usui H. *et al.*, *J Mol Neurosci.* 9(1): 55-60 (1997); Evans H.K. *et al.*, *Genomics* 77(1-2): 99-104 (2001)).

### OLFM2

[457] Probe set 223601 detects OLFM2 nucleic acid sequences. Expression of OLFM2 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	36.89	3.42	10	74.78	16.01	8	0.49	0.051	OLFM2

- B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [458] OLFM2 was also evaluated using real-time PCR. The results further show that OLFM2 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.34	0.029

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [459] Probe set 223601 detects OLFM2 nucleic acid sequences. Expression of OLFM2 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES			Fold Change	Students t test	Gene Name
Mean Expr	SEM	n	Mean Expr	SEM	n			
86.74	14.09	5	15.61	3.45	13	5.56	0.006	OLFM2

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [460] OLFM2 was over-expressed in 3T3-L1 adipocytes and the effect on  
5 basal and insulin stimulated glucose transport was determined

Glucose Transport Analysis in 3T3-L1 Adipocytes:

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (OLFM2/Con) +/- SEM (n=3)
	FC (OLFM2 / Con)	t test ( $\alpha$ )	FC (OLFM2 / Con)	t test ( $\alpha$ )	FC (OLFM2 / Con)	t test ( $\alpha$ )	
0	0.73	0.043	0.55	0.110	0.82	0.144	0.70±0.08
0.05	0.91	0.445	0.81	0.330	0.93	0.654	0.88±0.04
0.1	0.88	0.402	0.99	0.843	1.04	0.406	0.97±0.05
0.3	1.20	0.134	1.11	0.196	1.03	0.539	1.11±0.05
1	1.21	0.164	1.08	0.364	1.09	0.202	1.12±0.04
10	1.16	0.132	1.07	0.306	0.97	0.694	1.07±0.06

- Legend: "Con" indicates control 3T3-L1 adipocytes that do not express OLFM2. "FC" indicates the fold change defined as the following ratio; glucose transport in OLFM2-expressing 3T3-L1 adipocytes /glucose transport in non-OLFM2-expressing 3T3-L1 adipocytes. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.  
10

- [461] The results show that increasing the levels of OLFM2 in a cell such as  
15 an adipocyte leads to a corresponding decrease in glucose uptake. This indicates that decreasing the levels or activity of OLFM2 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

- [462] OLFM2 contains the following protein domains (designated with  
20 reference to SEQ ID NO:196): Signal peptide at amino acids 1 to 24; and Olfactomedin-like domain (PF02191) at amino acids 196 to 446. A soluble active secreted form of OLFM2 has been predicted and this is displayed in SEQ ID NO:197. OLFM2 is a protein which possess high similarity to olfactomedin 3 (rat Olfm3), which is known to interact with myocilin and is associated with glaucoma and disorders involving the anterior segment of the eye and the  
25 retina (Ortego J. *et al.*, *FEBS Lett.* 413(2): 349-53 (1997).



**OPN3**

[463] Probe set 219032 detects OPN3 nucleic acid sequences. Expression of OPN3 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	41.54	5.25	5	26.13	1.05	4	1.59	0.041	OPN3

- 5 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [464] OPN3 was also evaluated using real-time PCR. The results further show that OPN3 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	1.74	0.028

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [465] OPN3 contains the following protein domains (designated with reference to SEQ ID NO:203): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 58 to 309; and 7 transmembrane domains (TMHMM2.0) at amino acids 44 to 66, 78 to 100, 115 to 137, 158 to 180, 200 to 222, 258 to 280, 290 to 312. OPN3 is a member of the opsins receptor cluster of G protein-coupled receptors which may play a role in non-visual photic processes such as the entrainment of circadian rhythm or the regulation of pineal melatonin production (Blackshaw S., and Snyder S.H., *J Neurosci.* 19(10): 3681-90 (1999); Halford S. *et al.*, *Genomics* 72(2): 203-8 (2001)).

**PTPRE**

- [466] Probe set 221840 detects PTPRE nucleic acid sequences. Expression of PTPRE transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	92.02	8.84	10	50.7	4.22	8	1.81	0.001	PTPRE

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [467] PTPRE was also evaluated using real-time PCR. The results further show that PTPRE is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.8	0.001

[468] "Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR

- [469] PTPRE was over-expressed in 3T3-L1 adipocytes and the effect on basal and insulin stimulated glucose transport and Glut 4 translocation were determined.

Glut4 Translocation Analysis:

Insulin (nM)	Fold Change (Mean hPTPRE/Mean LacZ) (n=3)	t test (hPTPRE vs LacZ)
0	0.24	0.001
0.5	0.48	0.01
10	0.42	0.001

- Legend "Fold Change" indicates the following ratio; (Mean % of hPTPRE-expressing 3T3-L1 adipocytes that were scored positive for cell surface Glut4)/(Mean % of LacZ-expressing 3T3-L1 adipocytes that were scored positive for cell surface Glut4). "h" is human. "n" is the number of experiments.

[470] Increasing the levels of hPTPRE in 3T3-L1 adipocytes significantly inhibits basal and insulin-stimulated Glut4 translocation to the cell surface.

Glucose Transport Analysis in 3T3-L1 Adipocytes:

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (hPTPRE/Con) +/- SEM (n=3)
	FC (hPTPRE / Con)	t test ( $\alpha$ )	FC (hPTPRE / Con)	t test ( $\alpha$ )	FC (hPTPRE / Con)	t test ( $\alpha$ )	
0	1.06	0.747	0.85	0.554	0.84	0.021	0.92±0.07
0.05	0.80	0.095	1.23	0.163	0.66	0.046	0.90±0.17
0.1	0.67	0.208	0.73	0.041	0.58	0.048	0.66±0.04
0.3	0.82	0.115	0.73	0.027	0.59	0.040	0.71±0.07
1	0.94	0.340	0.76	0.006	0.65	0.018	0.78±0.08
10	0.95	0.192	0.83	0.067	0.72	0.008	0.83±0.07

**Legend:** "Con" indicates control 3T3-L1 adipocytes that do not express hPTPRE. "FC" indicates the fold change defined as the following ratio; glucose transport in hPTPRE-expressing 3T3-L1 adipocytes /glucose transport in non-PTPRE-expressing 3T3-L1 adipocytes. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

[471] The results show that increasing the levels of PTPRE in a cell such as an adipocyte leads to a corresponding decrease in glucose uptake. This indicates that decreasing the levels or activity of PTPRE in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[472] PTPRE contains the following protein domains (designated with reference to SEQ ID NO:213): Signal peptide at amino acids 1 to 19; Protein-tyrosine phosphatase (PF00102) at amino acids 159 to 393, 451 to 688; and 1 transmembrane domain (TMHMM2.0) at amino acids 47 to 69. PTPRE is found to exist in both a soluble, cytoplasmic and a transmembrane form. PTPRE is induced by IL1 and TNFA treatment in astrocytoma cells suggesting a role in the inflammatory response of the brain (Schumann, G. *et al.*, *Brain Res Mol Brain Res.* 62: 56-64 (1988)). Other studies suggest a role for PTPRE in RAS related signal transduction pathways, cytokines induced signaling, activation of voltage-gated K<sup>+</sup> channels and vascular development and angiogenesis (Tiran, Z.J. *et al.*, *J. Biol Chem.* 278: 17509-14(2003); Toledano-Katchalski, H., *et al.*, *Mol Cancer Res.* 1: 541-50 (2003); Thompson, L.J., *et al.*, *Am J Physiol Heart Circ Physiol.* 281: H396-403 (2001)).

[473] PTPRE can dephosphorylate the tyrosine phosphorylated insulin receptor (*see, e.g.*, Nakagawa Y. *et al. Zoolog Sci.* 22:169-75 (2005)). Thus, an exemplary assay for inhibitors of PTPRE comprises determining the ability of candidate compounds to inhibit the ability of purified PTPRE to dephosphorylate a tyrosine phosphorylated peptide,

e.g., a tyrosine phosphorylated peptide derived from the sequence of the insulin receptor and containing autophosphorylation sites such as Tyrosines 11158, 1162 and 1163.

### RDC1

- [474] Probe set 212977 detects RDC1 nucleic acid sequences. Expression of RDC1 transcripts was decreased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	169.88	9.84	10	238.53	13.88	8	0.71	0.001	RDC1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of diabetics in comparison to lean patients.

- [475] RDC1 was also evaluated using real-time PCR. The results further show that RDC1 is significantly under-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	0.65	0.008

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [476] Probe set 212977 detects RDC1 nucleic acid sequences. Expression of RDC1 transcripts was decreased in patients with insulin resistance compared to normal patients in the gene profiling experiment.

B/C	CORRELATION RD			Gene Name
	n	Corr Co-efficient	Students t test	
C	26	0.649	<0.005	RDC1

- B/C indicates sample is from Basal or Clamp; "Corr Co-efficient" indicates the relationship between glucose disposal rate (Rd) and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance; "n" indicates number of patient samples.

- [477] RDC1 was also evaluated using real-time PCR. The results further show that RDC1 is significantly decreased in patients with insulin resistance compared to normal patients.

Comparison	Expression Fold Change	t test
Correlation to Rd (26)	0.567	<0.005

"Corr Co-efficient" indicates the relationship between Rd and signal intensities. A positive co-efficient indicates down regulation whereas a negative co-efficient indicates up regulation of the gene with increasing insulin resistance. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- 5 [478] Probe set 212977 detects RDC1 nucleic acid sequences. Expression of RDC1 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

	OBESE			LEAN					
B/C	Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
B	237.66	36.98	5	153.85	5.4	4	1.54	0.086	RDC1

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

- 10 Change" indicates fold change of obese in comparison to lean patients.

[479] Probe set 212977 detects RDC1 nucleic acid sequences. Expression of RDC1 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
261.58	37.67	5	46.15	9.41	13	5.67	0.004	RDC1

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

- 15 indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [480] RDC1 contains the following protein domains (designated with reference to SEQ ID NO:223): 7 transmembrane receptor (rhodopsin family) (PF00001) at amino acids 61 to 315; and 7 transmembrane domains (TMHMM2.0) at amino acids 47 to 69, 82 to 104, 119 to 140, 160 to 182, 214 to 236, 255 to 277, 297 to 319. RDC1 is considered to be a new member of the rhodopsin family of G-protein coupled receptors. The protein is a co-receptor for human immunodeficiency viruses (HIV). Translocations involving this gene and HMGA2 on chromosome 12 have been observed in lipomas (Broberg, K., *et al.*, *Int J Oncol.* 21: 321-6 (2002); Shimizu, N., *et al.*, *J Virol.* 74: 619-26 (2000)).
- 20

**SLIT2**

[481] Probe set 209897 detects SLIT2 nucleic acid sequences. Expression of SLIT2 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	81.28	5.02	5	48.53	3.08	4	1.68	0.001	SLIT2

- 5 B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [482] SLIT2 was also evaluated using real-time PCR. The results further show that SLIT2 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	3.31	0.014

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [483] Cellular levels of SLIT2 were reduced in 3T3-L1 adipocytes using siRNA directed against SLIT2 and the effect on basal and insulin stimulated glucose transport was determined

**SLIT2 mRNA Level in 3T3-L1 Adipocytes Transfected with siRNA Oligonucleotides**

	Expt1	Expt2	Expt3	Mean FC ± SEM (n=3)
FC (siRNA/Scr)	0.29	0.22	0.26	0.26 ± 0.02

- 20 Legend: "siRNA" indicates Dharmacon Smartpool siRNA oligonucleotides directed against murine SLIT2. "Scr" indicates the Dharmacon Scramble siRNA Control. "FC" indicates the fold change defined as the following ratio; Level of SLIT2 mRNA in SLIT2 siRNA transfected 3T3-L1 adipocytes/Level of SLIT2 mRNA in Scramble siRNA transfected 3T3-L1 adipocytes. "n" is the number of experiments. SEM is the standard error of the mean.

**Glucose Transport in 3T3-L1 Adipocytes Transfected with siRNA Oligonucleotides**

Insulin (nM)	Expt 1		Expt 2		Expt 3		Mean FC (siRNA/Scr) ± SEM (n=3)
	FC (siRNA/Scr)	t-test	FC (siRNA/Scr)	t-test	FC (siRNA/Scr)	t-test	

0.0	1.235	0.020	0.948	0.001	0.639	0.600	0.941 ± 0.07
0.1	0.902	0.050	0.694	<0.001	0.573	0.020	0.723 ± 0.04
3.0	1.094	0.001	1.029	<0.001	0.709	0.490	0.944 ± 0.04

**Legend:** "siRNA" indicates Dharmacon Smartpool siRNA oligonucleotides directed against murine SLIT2. "Scr" indicates Dharmacon Scramble siRNA Control oligonucleotides. "FC" indicates the fold change defined as the following ratio; glucose transport in SLIT2 siRNA transfected 3T3-L1 adipocytes/glucose transport in Scramble siRNA transfected 3T3-L1 adipocytes. "n" is the number of experiments. SEM is the standard error of the mean.

[484] The results show that decreasing the levels of SLIT2 in a cell such as a adipocyte leads to a corresponding decrease in glucose uptake. This indicates that increasing the levels or activity of SLIT2 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[485] SLIT2 contains the following protein domains (designated with reference to SEQ ID NO:229): Signal peptide at amino acids 1 to 30; Leucine rich repeat C-terminal domain (PF01463) at amino acids 233 to 258, 454 to 479, 688 to 713, 883 to 908; Leucine rich repeat N-terminal domain (PF01462) at amino acids 27 to 54, 272 to 299, 505 to 532, 726 to 753; Leucine Rich Repeat (PF00560) at amino acids 56 to 79, 128 to 151, 176 to 199, 325 to 348, 349 to 372, 559 to 582, 607 to 630, 778 to 801, 802 to 825, 826 to 849; Laminin G domain (PF00054) at amino acids 1188 to 1319; and EGF-like domain (PF00008) at amino acids 922 to 954, 961 to 995, 1002 to 1033, 1040 to 1073, 1080 to 1111, 1125 to 1156, 1336 to 1367, 1375 to 1406, 1416 to 1447. A soluble active secreted form of SLIT2 has been detected (Nguyen Ba-Charvet, K.T. *et al.*, *J. Neurosci.*, 21: 4281-4289 (2001)) and this is displayed in SEQ ID NO:230. SLIT2 is the ligand for roundabout receptor ROBO1. Mammalian SLIT proteins may participate in the formation and maintenance of the nervous and endocrine systems by protein-protein interactions. SLIT2 has been reported to be a chemorepellant for neuronal migration in vivo induces branching of dorsal root ganglia axons (Nguyen Ba-Charvet K.T. *et al.*, *J Neurosci.*, 21(12): 4281-9 (2001); Nguyen Ba-Charvet K.T. *et al.*, *J Physiol Paris.* 96: 91-8 (2002)), SLIT2 also found to inhibit leukocyte chemotaxis induced by chemotactic factors (Wu, J.Y., *et al.*, *Nature* 410: 948-52 (2001)).

[486] [The human SLIT2 gene has been cloned and the proximal promoter has been identified (*see, e.g.*, Dallol A. *et al. Cancer Res.* 62:5874-80 (2002)). Therefore, one example of a method of screening for SLIT2 regulators is as follows. A SLIT2 promoter can be inserted upstream of a reporter gene such as  $\beta$ -galactosidase and expressed in cells.

Compounds that up-regulate the activity of the promoter may therefore be identified by measuring increased  $\beta$ -galactosidase activity.

### **TNFRSF21**

[487] Probe set 214581 detects TNFRSF21 nucleic acid sequences.

- 5 Expression of TNFRSF21 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	67.98	8.98	5	25.13	3.84	4	2.71	0.006	TNFRSF21

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- 10 [488] Probe set 214581 detects TNFRSF21 nucleic acid sequences.

Expression of TNFRSF21 transcripts was increased in rosi compared to vehicle treated cultures of primary human adipocytes in the gene profiling experiment.

B/C	ROSI			VEHICLE			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	52.77	3.6	3	33.8	2.99	3	1.56	0.016	TNFRSF21

B/C indicates sample is from Basal or Clamp; "Pre-Rosi" and "Post-Rosi" indicates sample was taken before or after 24 hours of rosiglitazone treatment; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of primary human adipocytes post-rosi in comparison to pre-rosi samples.

- 15 [489] TNFRSF21 was also evaluated using real-time PCR. The results further show that TNFRSF21 is significantly over-expressed in primary cultured human adipocytes treated with rosi when compared to vehicle.
- 20

Comparison	Expression Fold Change	t test
Post-Rosi (12)/ Pre-Rosi (12)	1.32	0.007

"Fold Change" indicates the fold expression calculated as the ratio of the mean rosi expression/ mean vehicle expression. Numbers in parentheses indicates the number of primary human adipocyte samples analyzed by real-time PCR.

[490] Probe set 214581 detects TNFRSF21 nucleic acid sequences.

- 25 Expression of TNFRSF21 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.



ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
70.58	15.04	5	16.12	2.78	13	4.38	0.021	TNFRSF21

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [491] TNFRSF21 was also evaluated using real-time PCR. The results further show that TNFRSF21 is significantly over-expressed in adipose tissues when compared to all other human adult tissues.

Comparison	Expression Fold Change	t test
Fat Tissues (5)/ All Other Tissues (13)	2.82	0.011

"Fold Change" indicates the fold expression calculated as the ratio of the mean adipose tissues expression/ mean other tissues expression. Numbers in parentheses indicates the number of human adult tissue samples analyzed by real-time PCR.

- [492] TNFRSF21 was over-expressed in 3T3-L1 adipocytes and the effect on basal and insulin stimulated glucose transport was determined.

#### Glucose Transport

Insulin (nM)	Experiment 1		Experiment 2		Experiment 3		Mean FC (hTNFRSF21/Con) +/- SEM (n=3)
	FC (hTNFRSF21/Con)	t test	FC (hTNFRSF21/Con)	t test	FC (hTNFRSF21/Con)	t test	
0	0.99	0.944	0.78	0.066	0.71	0.127	0.83±0.08
0.05	0.74	0.048	0.84	0.029	0.74	0.107	0.77±0.03
0.1	0.73	0.106	0.75	0.133	0.68	0.306	0.72±0.02
0.3	0.79	0.062	0.88	0.112	0.88	0.709	0.85±0.03
1	0.84	0.075	0.96	0.479	0.92	0.621	0.91±0.04
10	0.95	0.760	0.93	0.400	0.93	0.839	0.94±0.01

- Legend: "Con" indicates control 3T3-L1 adipocytes that do not express hTNFRSF21. "FC" indicates the fold change defined as the following ratio; glucose transport in hTNFRSF21-expressing cells/glucose transport in non-TNFRSF21-expressing cells. "h" is human. "n" is the number of experiments. SEM is the standard error of the mean.

- [493] The results show that increasing the levels of TNFRSF21 in a cell such as an adipocyte leads to a corresponding decrease in glucose uptake. This indicates that decreasing the levels or activity of TNFRSF21 in tissues of insulin resistant patients or diabetic patients will increase the ability of such tissues to take up glucose and hence, will provide an effective treatment for insulin resistance and diabetes.

[494] TNFRSF21 contains the following protein domains (designated with reference to SEQ ID NO:236): Signal peptide at amino acids 1 to 41; Death domain (PF00531) at amino acids 416 to 498; TNFR/NGFR cysteine-rich region (PF00020) at amino acids 50 to 88, 91 to 131, 133 to 168, 171 to 211; and 1 transmembrane domain (TMHMM2.0) at amino acids 350 to 369. TNFRSF21 has been shown to activate NF-kappaB and MAPK8/JNK, and induce cell apoptosis. Through its death domain, this receptor interacts with TRADD protein, which is known to serve as an adaptor that mediates signal transduction of TNF-receptors. Knockout studies in mice suggested that this gene plays a role in T-helper cell activation, and may be involved in inflammation and immune regulation (Pan G. *et al.*, *FEBS Lett.* 431(3): 351-356 (1998); Kasof G.M. *et al.*, *Oncogene* 20(55): 7965-7975 (2001)).

[495] TNFRSF21 is up-regulated by TNFalpha. which itself has been associated with states of insulin resistance (*see, e.g.*, Hotamisligil GS. *J Intern Med.* 245:621-625 (1999); Kasof GM *et al.* *Oncogene.* 20 :7965-7975 (2001)). An exemplary assay to identify compounds that suppress TNFRSF21 can thus employ cells such as LnCAP to screen candidate compounds for the ability to inhibit TNFalpha-induced up-regulation of TNFRSF21 mRNA or protein.

### TNFSF13B

[496] Probe set 223501 detects TNFSF13B nucleic acid sequences.

Expression of TNFSF13B transcripts was increased in diabetic compared to lean patients in the gene profiling experiment.

B/C	DIABETIC			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
C	44.86	3.06	10	28.79	3.14	8	1.56	0.002	TNFSF13B

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression;

"SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold

Change" indicates fold change of diabetics in comparison to lean patients.

[497] TNFSF13B was also evaluated using real-time PCR. The results further show that TNFSF13B is significantly over-expressed in subcutaneous adipose from diabetic individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Diabetic (10) / Lean (8)	1.54	0.007

"Fold Change" indicates the fold expression calculated as the ratio of the mean diabetic expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [498] TNFSF13B contains the following protein domains (designated with reference to SEQ ID NO:242): Signal anchor at amino acids 0 to 0; TNF(Tumour Necrosis Factor) family (PF00229) at amino acids 166 to 284; and 1 transmembrane domain (TMHMM2.0) at amino acids 48 to 70. A soluble active secreted form of TNFSF13B has been detected (Schneider, P. *et al.*, *J Exp. Med.*, 189: 1747-1756 (1999)) and this is displayed in SEQ ID NO:243. TNFSF13B is a ligand for multiple receptors including TNFRSF13B/TACI, TNFRSF17/BCMA, and TNFRSF13C/BAFFR. This cytokine is expressed in B cell lineage cells, and acts as a potent B cell activator. It has been also shown to play an important role in the proliferation and differentiation of B cells (Patke, A., *et al.*, *Curr Opin Immunol.* 16: 251-5 (2004); Schneider, P. and Tschopp, J. *Immunol Lett.* 88: 57-62 (2003)).

## 15 TNFSF14

[499] Probe set 207907 detects TNFSF14 nucleic acid sequences. Expression of TNFSF14 transcripts was increased in omental tissue compared to all other human adult tissues in the gene profiling experiment.

OMENTAL TISSUE			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
21.2	nd	1	2.71	0.66	17	7.83	n.d.	TNFSF14

- "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of tissue samples; "Fold Change" indicates fold change of omental tissues in comparison to all other human adult tissues profiled.

[500] TNFSF14 was also evaluated using real-time PCR. The results further show that TNFSF14 is significantly over-expressed in omental tissue when compared to all other human adult tissues.

Comparison	Expression Fold Change	t test
Omental (1) / All Other Tissues (17)	16.75	n.d.

- "Fold Change" indicates the fold expression calculated as the ratio of the mean omental tissue expression/ mean other tissues expression. Numbers in parentheses indicates the number of human adult tissue samples analyzed by real-time PCR.

[501] TNFSF14 contains the following protein domains (designated with reference to SEQ ID NO:251): Signal anchor at amino acids 0 to 0; TNF(Tumour Necrosis Factor) family (PF00229) at amino acids 93 to 240; and 1 transmembrane domain (TMHMM2.0) at amino acids 36 to 58. A soluble active secreted form of TNFSF14 has been detected (Harrop, J.A., *et al.*, *J Biol Chem.* 273:27548-56 (1998)) and this is displayed in SEQ ID NO:252. TNFSF14 is the ligand for HVEM (TNFRSF14) and is reported to induce lymphocyte proliferation induces apoptosis and suppresses in vivo tumor formation and facilitate herpes virus entry (Mauri D.N. *et al.*, *Immunity* 8(1): 21-30 (1998); Zhai Y. *et al.*, *J Clin Invest.* 102(6): 1142-51 (1998); Castellano, R. *et al.*, *J Biol Chem.* 277(45): 42841-51 (2001)).

### TPSB2

[502] Probe set 205683 detects TPSB2 nucleic acid sequences. Expression of TPSB2 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	154.7	20.57	5	69.33	8.15	4	2.23	0.011	TPSB2

B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

[503] TPSB2 was also evaluated using real-time PCR. The results further show that TPSB2 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	2.7	0.063

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

[504] Probe set 205683 detects TPSB2 nucleic acid sequences. Expression of TPSB2 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
174.74	30.6	5	48.31	13.35	13	3.62	0.01	TPSB2

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [505] TPSB2 contains the following protein domains (designated with reference to SEQ ID NO:262): Signal peptide at amino acids 1 to 18; and Trypsin (PF00089) at amino acids 31 to 267. A soluble active secreted form of TPSB2 has been detected and this is displayed in SEQ ID NO:263. TPSB2 is a tryptase beta 2 and belongs to the family of mast cell serine proteases which have been implicated as mediators in the pathogenesis of asthma and other allergic and inflammatory disorders. Beta tryptases appear to be the main isoenzymes expressed in mast cells (Pallaoro M. *et al.*, *J Biol Chem.*, 274(6): 3355-62 (1999).

### WISP2

[506] Probe set 205792 detects WISP2 nucleic acid sequences. Expression of WISP2 transcripts was increased in obese compared to lean patients in the gene profiling experiment.

B/C	OBESE			LEAN			Fold Change	Students t test	Gene Name
	Mean Expr	SEM	n	Mean Expr	SEM	n			
B	149.48	14.39	5	97.75	11.6	4	1.53	0.044	WISP2

- B/C indicates sample is from Basal or Clamp; "Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n" indicates number of patient samples; "Fold Change" indicates fold change of obese in comparison to lean patients.

- [507] WISP2 was also evaluated using real-time PCR. The results further show that WISP2 is significantly over-expressed in subcutaneous adipose from obese individuals when compared to subcutaneous adipose from lean individuals.

Comparison	Expression Fold Change	t test
Obese (5) / Lean (4)	2.24	0.002

"Fold Change" indicates the fold expression calculated as the ratio of the mean obese expression/ mean lean expression. Numbers in parentheses indicates the number of patient samples analyzed by real-time PCR.

- [508] Probe set 205792 detects WISP2 nucleic acid sequences. Expression of WISP2 transcripts was increased in adipose tissues compared to all other human adult tissues in the gene profiling experiment.

ADIPOSE TISSUES			OTHER TISSUES					
Mean Expr	SEM	n	Mean Expr	SEM	n	Fold Change	Students t test	Gene Name
167.08	43.22	5	26.37	9.83	13	6.34	0.029	WISP2

"Mean Expr" indicates mean expression; "SEM" indicates standard error of mean; "n"

indicates number of tissue samples; "Fold Change" indicates fold change of adipose tissues in comparison to all other human adult tissues profiled.

- [509] WISP2 contains the following protein domains (designated with reference to SEQ ID NO:269): Signal peptide at amino acids 1 to 23; Insulin-like growth factor binding protein (PF00219) at amino acids 26 to 96; von Willebrand factor type C domain (PF00093) at amino acids 100 to 163; Thrombospondin type 1 domain (PF00090) at amino acids 196 to 237; and TNFR/NGFR cysteine-rich region (PF00020) at amino acids 39 to 70. A soluble active secreted form of WISP2 has been detected (Pennica, D., *et al.*, *Proc Natl Acad Sci U S A.* 95:14717-22 (1998)) and this is displayed in SEQ ID NO:270. WISP2 is a member of the CCN family of growth factors and is found to promote the adhesion of osteoblast cells. Decreased expression of WISP2 may be therapeutic in treatment of breast cancer (Kumar S. *et al.*, *J Biol Chem.*, 274(24): 17123-31 (1999); Pennica D. *et al.*, *Proc Natl Acad Sci. USA.* 95(25): 14717-22 (1998)). WISP2 is also found to inhibit vascular smooth muscle cell proliferation, motility, and invasiveness. (Lake, A.C. *et al.*, *Am J Pathol.* 162: 219-31 (2003).

## SEQUENCE LISTING

SEQ ID NO: 1

gi|18390318|ref|NM\_015419.1| Homo sapiens adlcan (DKFZp564I1922), mRNA

```

5   1 atgccaagc gcgcgactg gggggccctc tccgtggtgc tgatcctgct ttggggccat
    61 cgcgagtg cgcctggcctg cccgcaccc tctgcctgct acgtcccccag cgaggtccac
    121 tgcacgttcc gatccctggc ttcctgccc gctggcattg ctgacacagt ggaaagaatc
    181 aatttggggg ttaatatgcat acaggccctg tcagaaacct catttgagg actgaccaag
    241 ttggagctac ttatgattca cggcaatgag atcccaagca tccccgatgg agctttaaga
    301 gacctcagct ctcttcagg tttcaagttc agctacaaca agctgagagt gatcacagga
10  361 cagaccctcc aggtgtcttc taacttaatg aggtgcaca ttgaccaca caagatcgag
    421 tttatccacc ctcaagcttt caacggctta acgtctctga ggctactcca tttggaagga
    481 aatctcctcc accagctgca cccagcacc tctccacgt tcacattttt ggattatttc
    541 agactctcca ccataaggca cctctactta gcagagaaca tggtagaac tcttctgccc
    601 agcatgcttc ggaacatgcc gcttctggag aatctttact tgcagggaaa tcttggacc
15  661 tgcgattgtg agatgagatg gtttttggaa tgggatgcaa aatccagagg aattctgaag
    721 tgtaaaaagg acaaagctta tgaaggcggt cagttgtgtg caatgtgctt cagtccaaag
    781 aagttgtaca aacatgagat acacaagctg aaggacatga cttgtctgaa gccttcaata
    841 gagtccccctc tgagacagaa caggagcagg agtattgagg aggagcaaga acaggagag
    901 gatggtgcca gccagctcat cctggagaaa ttccaactgc cccagtggag catctcttg
20  961 aatatgaccg acgagcacgg gaacatgggt aacttggctc gtgacatcaa gaaaccaatg
1021 gatgtgtaca agattcactt gaaccaaacg gatcctccag atattgacat aaatgcaaca
1081 gttgccttgg actttgagtg tccaatgacc cgagaaaact atgaaaagct atggaattg
1141 atagcatact acagtgaagt tcccgtgaag ctacacagag agctcatgct cagcaaagac
1201 cccagagtca gctaccagta caggcaggat gctgatgagg aagctcttta ctacacagg
25  1261 gtgagagccc agattcttgc agaaccagaa tgggtcatgc agccatccat agatattccag
1321 ctgaaccgac gtcagagtac ggccaagaag gtgctacttt cctactacac ccagtattct
1381 caaacaatat ccacaaaga tacaaggcag gctcggggca gaagctgggt aatgattgag
1441 cctagtggag ctgtgcaaa agatcagact gtccctggaag ggggtccatg aaatgcagc
1501 tgcaacgtga aagcttctga gagtccatct atcttctggg tgctccaga tgggtccatc
30  1561 ctgaaagcgc ccatggatga cccagacagc aagttctcca ttctcagcag tgggtggctg
1621 aggatcaagt ccatggagcc atctgactca ggcttgtagc agtgcatgct tcaagtgagg
1681 gatgaaatgg accgcatggt atatagggtg cttgtgcagt ctccctccac tcagccagcc
1741 gagaaagaca cagtgacaat tggcaagaac ccaggggagt cggtgacatt gccttgcaat
1801 gctttagcaa taccgaagc ccaccttagc tggattcttc caaacagaag gataattaat
35  1861 gatttggcta acacatcaca gtatatactg ttgccaatg gaactctttc gatcccaag
1921 gtccaagtca gtgatagtgg ttactacaga tgtgtggctg tcaaccagca agggcgagac
1981 cattttacgg tgggaatcac agtgaccaag aaagggctctg gcttgccatc caaaagaggc
2041 agacgcccag gtgcaaaggc tctttccaga gtcagagaag acatcgtgga ggatgaaggg
2101 ggctcgggca tgggagatga agagaacact tcaaggagac ttctgcatcc aaaggaccaa
40  2161 gaggtgttcc tcaaaacaaa ggatgatgcc atcaatggag acaagaaagc caagaaaggg
2221 agaagaaagc tgaactctg gaagcattcg gaaaaagaac cagagaccaa tgttcagaa
2281 ggtcgagag tgtttgaatc tagacgaagg ataaacatgg caaacaacaa gattaatccg
2341 gagcgctggg ctgatatttt agccaaagtc cgtgggaaaa atctccctaa gggcacagaa
2401 gtacccccgt tgattaaaac cacaagtcct ccctccttga gcctagaagt cacaccacct
45  2461 tttcctgctg tttctcccc ctcagcatct cctgtgcaga cagtaaccag tgcgaagaa
2521 tctcagcag atgtacctct acttggtgaa gaagagcacg ttttgggtac catttctca
2581 gccagcatgg ggctagaaca caaccacaat ggagttattc ttgttgaaac tgaagtaaca
2641 agcacacctc tggaggaagt tgttgatgac ctttctgaga agactgagga gataacttcc
2701 actgaaggag acctgaaggc gacagagcc cctacactta tatctgagcc ttatgaacca
50  2761 tctcctactc tgcacacatt agacacagtc tatgaaaagc ccacccatga agagacggca
2821 acagaggggtt ggtctgcagc agatgttgga tcgtcaccag agcccacatc cagttagtat
2881 gagcctccat tggatgctgt ctccttggtc gagtctgagc ccatgcaata ctttgaccca
2941 gatttggaga ctaagtcaca accagatgag gataagatga aagaagacac ctttgacac
3001 cttactccaa cccccacat ctgggttaat gactccagta catcacagtt atttgaggat
55  3061 tctactatag ggaaccagg tgtcccaggc caatcacatc tacaaggact gacagacaac
3121 atccaccttg tgaaaagtag tctaagcact caagacacct tactgattaa aaagggatat

```

	3181	aaagagatgt	ctcagacact	acagggagga	aatatgctag	agggagaccc	cacacactcc
	3241	agaagtcttg	agagtggagg	ccaagagagc	aaatccatca	ctttgcctga	ctccacactg
	3301	ggtataatga	gcagtatgtc	tccagttaag	aagcctgcgg	aaaccacagt	tggtaccctc
5	3361	ctagacaaag	acaccacaac	agtaacaaca	acaccaaggc	aaaaagtgtc	tccgtcatcc
	3421	accatgagca	ctcacccttc	tcgaaggaga	cccaacggga	gaaggagatt	acgccccaac
	3481	aaattccgcc	accggcacaa	gcaaacccca	cccacaactt	ttgccccatc	agagactttt
	3541	tctactcaac	caactcaagc	acctgacatt	aagatttcaa	gtcaagtgga	gagttctctg
	3601	gttcctacag	cttgggtgga	taacacagtt	aataccccca	aacagttgga	aatggagaag
10	3661	aatgcagAAC	ccacatccaa	gggaacacca	cggagaaaac	acgggaagag	gccaacacaa
	3721	catcgatata	cccccttctac	agtgagctca	agagcgtccg	gatccaagcc	cagcccttct
	3781	ccagaaaaa	aacatagaaa	cattgtttact	cccagttcag	aaactatact	tttgccctaga
	3841	actgtttctc	tgaaaactga	gggcctttat	gattccttag	attacatgac	aaccaccaga
	3901	aaaatatatt	catcttacc	taaagtccaa	gagacacttc	cagtcacata	taaacccaca
	3961	tcagatggaa	aagaaattaa	ggatgatgtt	gccacaaatg	ttgacaaaac	taaaagtgc
15	4021	atttttagtca	ctgggtgaatc	aattactaat	gccataccaa	cttctcgctc	cttgggtctcc
	4081	actatgggag	aattttaagg	agaatcctct	cctgtaggct	ttccaggaac	tccaacctgg
	4141	aatccctcaa	ggacggccca	gcctggggag	ctacagacag	acatacctgt	taccacttct
	4201	ggggaaaaatc	ttacagaccc	tccccttctt	aaagagcttg	aggatgtgga	tttcaacttcc
	4261	gagtttttgt	cctctttgac	agtctccaca	ccatttcacc	aggaagaagc	tggttcttcc
20	4321	acaactctct	caagcataaa	agtggaggtg	gcttcaagtc	aggcagaaac	caccaccctt
	4381	gatcaagatc	atcttgaaac	cactgtggct	attctccttt	ctgaaactag	accacagaat
	4441	cacaccctta	ctgctgccc	gatgaaggag	ccagcatcct	cgtccccatc	cacaattctc
	4501	atgtctttgg	gacaaaccac	caccactaag	ccagcacttc	ccagtccaag	aatatctcaa
	4561	gcattctagag	attccaagga	aaatgttttc	ttgaattatg	tggggaatcc	agaaacagaa
25	4621	gcaacccag	tcaacaatga	aggaacacag	catatgtcag	ggccaaatga	attatcaaca
	4681	ccctcttccg	accgggatgc	atttaacttg	tctacaaagc	tggaattgga	aaagcaagta
	4741	tttggtagta	ggagtctacc	acgtggccca	gatagccaac	gccaggatgg	aagagttcat
	4801	gcttctcatc	aactaaccag	agtccctgcc	aaacccatcc	taccaacagc	aacagtgagg
	4861	ctacctgaaa	tgctccacaca	aagcgtctcc	agatactttg	taacttccca	gtcacctcgt
30	4921	cactggacca	acaaaccgga	aataactaca	tatccttctg	gggctttgcc	agagaacaaa
	4981	cagtttaca	ctccaagatt	atcaagtaca	acaattcctc	tcccattgca	catgtccaaa
	5041	cccagcatte	ctagtaagtt	tactgaccga	agaactgacc	aattcaatgg	ttactccaaa
	5101	gtgtttggaa	ataacaacat	ccctgaggca	agaaaccag	ttggaaagcc	tcccagtcga
	5161	agaattctct	attattccaa	tggaagactc	cctttcttta	ccaacaagac	tctttctttt
35	5221	ccacagttgg	gagtcacccg	gagaccccag	ataccacttt	ctcctgcccc	agtaatgaga
	5281	gagagaaaag	ttattccagg	ttcctacaac	aggatacatt	cccatagcac	cttccatctg
	5341	gactttggcc	ctccggcacc	tccgttgttg	cacactccgc	agaccacggg	atcaccctca
	5401	actaacttac	agaatatccc	tatggtctct	tccaccacga	gttctatctc	ctttataaca
	5461	tcttctgtcc	agtccctcag	aagcttccac	cagagcagct	caaagtctct	tgacaggagga
40	5521	cctcctgcat	ccaaattctg	gtctcttgag	gaaaagcccc	aaatcctcac	caagtcacca
	5581	cagactgtgt	ccgtcaccgc	tgagacagac	actgtgttcc	cctgtgaggc	aacaggaaaa
	5641	ccaaagcctt	tcgttacttg	gacaaagggt	tccacaggag	ctcttatgac	tccgaatacc
	5701	aggatacaac	ggtttgagg	tctcaagaac	ggtaccttag	tgatacggaa	ggttcaagta
	5761	caagatcgag	gccagtatat	gtgcaccgcc	agcaacctgc	acggcctgga	caggatgggtg
45	5821	gtcttgcttt	cggtcaccgt	gcagcaacct	caaatcctag	cctcccaacta	ccaggacgtc
	5881	actgtctacc	tgggagacac	cattgcaatg	gagtgctctg	ccaaagggac	cccagcccc
	5941	caaatcttct	ggatcttccc	tgacaggagg	gtgtggcaaa	ctgtgtcccc	cgtggagagc
	6001	cgcatcacc	tgacgaaaaa	cggaccctt	tccatcaagg	aggcgtcctt	ctcagacaga
50	6061	ggcgtctata	agtgcgtggc	cagcaatgca	gccggggcgg	acagcctggc	catccgctg
	6121	cacgtggcgg	cactgcccc	cgttatccac	caggagaagc	tggaagaacat	ctcgtgccc
	6181	ccggggctca	gcattcacat	tactgcaact	gccaaggctg	cgccctgccc	cagcgtgcgc
	6241	tgggtgctcg	gggacgggtac	ccagatccgc	ccctcgaggt	tcctccacgg	gaacttgttt
	6301	gttttcccca	acgggacgct	ctacatccgc	aacctcgccg	ccaaggacag	cgggcgctat
	6361	gagtgctgg	ccgccaacct	ggtaggctcc	gcgcgcagga	cgggtgcagct	gaacgtgcag
55	6421	cgtgcagcag	ccaacgcgcg	ctacacggcg	acctcccgcc	ggaggacgga	cgtcaggtac
	6481	ggaggaaccc	tcaagctgga	ctgcagcgcc	tcgggggacc	cctggccgcg	catcctctgg
	6541	aggctgccgt	ccaagaggat	gatcgacgcg	ctcttcagtt	ttgatagcag	aatcaagggtg
	6601	tttgccaatg	ggaccctgg	ggtgaaatca	gtgacggaca	aagatgccgg	agattacctg
	6661	tgcgtagctc	gaaataaggt	tggtgatgac	tacgtggtgc	tcaaagtggga	tggtggtgatg
60	6721	aaaccggcca	agattgaaca	caaggaggag	aacgaccaca	aagtcttcta	cgggggtgac
	6781	ctgaaagtgg	actgtgtggc	caccgggctt	cccaatcccc	agatctcctg	gagcctccca



```

5  6841 gacgggagtc tggatgaactc cttcatgcag tccgatgaca gcgggtggacg caccaagcgc
6901 tatgtcgtct tcaacaatgg gacactctac tttaacgaag tggggatgag ggaggaagga
6961 gactacacct gctttgctga aaatcaggtc gggaaggacg agatgagagt cagagtcaag
7021 gtggtgacag cgcccgcac catccggaac aagacttact tggcggttca ggtgcctat
7081 ggagacgtgg tcaactgtagc ctgtgaggcc aaaggagaac ccatgcccaa ggtgacttgg
7141 ttgtcccca ccaacaaggat gatccccacc tcctctgaga agtatcagat ataccaagat
7201 ggcactctcc ttattcagaa agcccagcgt tctgacagcg gcaactacac ctgcctggtc
7261 aggaacagcg cgggagagga taggaagacg gtgtggattc acgtcaacgt ccagccaccc
7321 aagatcaacg gtaaccccaa ccccatcacc accgtgcggg agatagcagc cgggggcagt
10 7381 cggaaactga ttgactgcaa agctgaaggc atccccaccc cgagggtgtt atgggctttt
7441 cccgaggggtg tgggttctgcc agctccatac tatggaaacc ggatcactgt ccatggcaac
7501 ggttcctgtg acatcaggag tttgaggaa agcgactccg tccagctggg atgcatggca
7561 cgcaacgagg gaggggaggc gagggttgac gtgcagctca ctgtcctgga gccatggag
7621 aaacccatct tccacgaccc gatcagcgag aagatcacgg ccatggcggg ccacaccatc
15 7681 agcctcaact gctctgccgc ggggaccccg acaccagcc tgggtgtgggt ccttcccaat
7741 ggcaccgacg tgcagagtgg acagcagctg cagcgttct accacaaggc tgacggcatg
7801 ctacacatta gcggtctctc ctccgtggac gctggggcct accgctgcgt ggcccgcagt
7861 gccgctggcc acacggagag gctggtctcc ctgaagggtg gactgaagcc agaagcaaac
7921 aagcagatc ataactggt cagcatcac aatggtgaga cctgaagct cccctgcacc
20 7981 cctcccgggg ctggggcagg acgtttctcc tggacgctcc ccaatggcat gccatggag
8041 ggcccccaaa cctgggacg cgtttctctt ctggacaatg gcaccctcac ggttcgtgag
8101 gcctcggtgt ttgacagggg tacctatgta tgcaggatgg agacggagta cggcccttcg
8161 gtcaccagca tccccgtgat tgtgatcgcc tatcctcccc ggatcaccag cgagcccacc
8221 ccggtcatct acacccggcc cgggaacacc gtgaaactga actgcatggc tatggggatt
25 8281 cccaaagctg acatcacgtg ggagttaccg gataagtcgc atctgaaggc aggggttcag
8341 gctcgtctgt atggaacag atttctcac cccagggat cactgacct ccagcatgcc
8401 acacagagag atgccggtt ctacaagtgc atggcaaaaa acattctcgg cagtgcctcc
8461 aaaacaactt acatccacgt cttctgaaat gtggattcca gaatgattgc ttaggaactg
8521 acaacaagc ggggtttgta agggaagcca ggttggggaa taggagctct taaataatgt
30 8581 gtcacagtgc atggtggcct ctggtgggtt tcaagttag gttgatcttg atctacaatt
8641 gttgggaaaa ggaagcaatg cagacacgag aaggagggtc cagccttgct gagacacttt
8701 cttttgtgtt tacatcatgc caggggcttc attcagggtg tctgtgctct gactgcaatt
8761 tttctctttt tgcaaatgcc actcgaactg cttcataagc gtccatagga tatctgagga
8821 acattcatca aaaataagcc atagacatga acaacacctc actaccccat tgaagacga
35 8881 tcacctagtt aacctgctgc agtttttaca tgatagactt tgttccagat tgacaagtca
8941 tctttcagtt atttctctg tcaattcaaa actccagctt gcccataaag gatttagaac
9001 cagagtgact gatatatata tatatatatt aattcagagt tacatacata cagctaccat
9061 tttatatgaa aaaagaaaaa catttcttcc tggaaactcac tttttatata atgttttata
40 9121 tatatatatt ttcctttcaa atcagacgat gagactagaa ggagaaatac tttctgtctt
9181 attaaaatta ataaattatt ggtctttaca agacttggat acattacagc agacatggaa
9241 atataatttt aaaaaatttc tctccaacct ccttcaaatt cagtcaccac tgttatatta
9301 ccttctccag gaaccctcca gtggggaagg ctgcgatatt agatttctt gtatgcaag
9361 tttttgttga aagctgtgct cagaggaggt gagaggagag gaaggagaaa actgcatcat
9421 aactttacag aattgaatct agagtcttcc ccgaaaagcc cagaaacttc tctgcagtat
45 9481 ctggcttgtc catctgggtc aaggtggctg cttcttcccc agccatgagt cagtttgtgc
9541 ccatgaataa tacacgacct gttatttcca tgactgcttt actgtatttt taaggtaaat
9601 atactgtaca ttgataata aaataatatt ctcccaaaaa aaaaa

```

SEQ ID NO: 2

Amino acid sequence of human ADL1CAN encoded by the DNA sequence shown in SEQ ID NO: 1.

```

50  MPKRAHWGALS VVLLWGHPRVALACPHPCACYPVSEVHCTFRSLASVPAGIARHVERI
    NLGFNSIQALSETSFAGLTKELLMIHGNEIPSPDGLRDLSSLQVFKFSYNKLRVITG
    QTLQGLSNLMRLHIDHNKIEFIHQAFNGLTSLRLLHLEGNLLHQLHPSTFSTFTFLDYF
    RLSTIRHLYLAENMVRTLPAASMLRNMPLEENLYLQGNPWTCCEMRWFLEWDAKSRGILK
55  CKKDKAYEGGQLCAMCFSPKKLYKHEIHKLKDMTCLKPSIESPLRQNRSRSEEEQEQUEE
    DGGSQLLEKFLPQWSISLNMTEHGNMNVNLVCDIKKPMVYKIHNLQTDPPDIDINAT
    VALDFECPMTRNEYKWLKLIAYYSEVPVKLHRELMLSKDPRVSYQYRQDADEEALYYTG

```

VRAQILAEPEWVMQPSIDIQLNRRQSTAKKVLLSYTQYSQTIISTKDTRQARGRSWVMIE  
 PSGAVQRDQTVLEGGPCQLSCNVKASESPSIFWVLPDGSILKAPMDDPDSPKFSILSSGWL  
 RIKSMEPSDSGLYQCIAQVRDEMDRMVYRVLVQSPSTQPAEKDVTVTIGKNPGESVTLPCN  
 ALAIPHAHLISWILPNRRIINDLANTSHVYMLPNGTSLIPKVQVSDSGYYRCVAVNQGGAD  
 5 HFTVGITVTKKGSGLPKRRRPGAKALSRVREDIVEDEGGSGMGDEENTSRRLLHPKDQ  
 EVFLKTKODAINGDRKAKKGRRLKLWKHSEKEPETNVAEGRRVFESRRRINMANKQINP  
 ERWADILAKVRGKNLPKGTEVPPLIKTTSPPSLSEVTPPPPAVSPPSASPVQTVTSABE  
 SSADVPLLGEHHVLTGSISSASMGLEHNHNGVILVEPEVTSTPLEEVVDDLSEKTEBIS  
 10 TEGDLKGTAAPTILISEPYEPSPTLHTLDTVYEKPTHEETATEGWSAADVGSSPEPTSSEY  
 EPPLDAVSLAESEPMQYFDPDLETKSQPDEDKMKEDTFAHLTPTPTIWNDSSTSOLFED  
 STIGEPGVPGQSHLQGLTDNIHLVSSSLSTQDILLIKGMKEMSQTLOGGNMLEGDPHTS  
 RSSESEGGESKSIITLDPSTLGIMSSMSPVKKPAETTGTLLDKDTTTTVTTPRQKVAPSS  
 TMSHTPSRRRPNRRRLRPKNKFRHRHKQTPPTTFAPSETFSTQPTQAPDIKISSQVLESS  
 VPTAWVDNTVNTPKQLEMEKNAEPTSKGTPRRKHGKRPKNKHRYTPSTVSSRASGSKSPS  
 15 PENKHRNIVTPSSETILLPRTVSLKTEGPDYSLDYMTTRKIYSSYPKVQETLPVTYKPT  
 SDGKEIKDDVATNVDKHKSILVTGESITNAIPTSRSLVSTMGFEKEESSPVGFPGTPTW  
 NPSRTAQPGRLQTDIPVTTSGENLTDPLLKELEDVDFTEFLSSLTVSTPFHQEEAGSS  
 TTLSSIKVEVASSQAETTTLDQDHELTVAILLSETRPQNHTPTAARMKEPASSSPSTIL  
 MSLGQTTTTTKPALSPRISQASRDSKENVFLNYVGNPTEATPVNNEGTHMSGPNELST  
 20 PSSDRQTTNKLSTKLELEKQVFGSRSLPRGPDSDRGVHASHQLTRVPAKPIILPTATVR  
 LPEMSTQSASRYFVTSQSPRHWTNKPEITTYPSGALPENKQFTTTPRLSSTTIPLPLHMSK  
 PSIPSKFTDRRTDQFNGYSKVFQNNNIPARNPVGKPPSPRIPHYSGRLPFFTNKTLSP  
 PQLGVTRRPQIPTSPAPVMRERKVI PGSYNRIHSHSTFHLDFGPPAPPLHTPQTGSPS  
 TNLQNI PMVSSSTQSSISFITSSVQSSGSHQSSSKFFAGGPPASKFWSLGEKPKIILTKSP  
 25 QTVSVTAETDVFPCATGKPKPFVWTWKVSTGALMTPNTRIQRFEVLKNGTLVIRKVQV  
 QDRGQYMCATASNLHGLDRMVVLLSVTVQQPQILASHYQDVTVYLGDTIAMECLAKGTPAP  
 QISLWIFPDRRVNQTCLVRNSAGEDRKTVWIHVNQPPKINGNPNPITTVREIAAGGS  
 HVAALPPVIHQEKLLENISLPPGLSIHICTAKAAPLPSVRWVLGDGTQIRPSQFLHGNLF  
 VFPNGTLYIRNLAPKDSGRYECVAANLVGSARRTVQLNVQRAAANARITGTSRRRTDVRY  
 30 GGTLLKDCSASGDPWPRLWRLPSKRMIDALFSFDSRIKVFANGTLVVKSVTDKADGDL  
 CVARNKVGDDYVVLKVDVVMKPAKIEHKEENDHKVFYGGDLKVDCAVATGLPNPEISWSLP  
 DGLSVNSFMQSDSDSGRTRKRYVFNNGTLYFNEVGMREEGDYTCFAENQVGKDEMRVRK  
 VVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVWLSPTNKVIPTSSEKYQIYQD  
 GTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNQPPKINGNPNPITTVREIAAGGS  
 35 RKLIDCKAEGIPTPRVLWAFPEGVLPAPYVGNRITVHGNGSLDIRSLRKSQVQLVCM  
 RNEGGEARLIVQLTVLEPMKPIPHDPISEKITAMAGHTISLNCSAAGTPTPSLVWVLPN  
 GTDLQSGGQLQRFYHKADGMLHISGLSSVDAGAYRCVARNAAGHTERLVSLKVGKPEAN  
 KQYHNLVSIINGETLKLPCPTPPGAGQGRFSWTL PNGMHLEGPQTLGRVSLLDNGTLTVRE  
 ASVFDRTGYVCRMETEGPSVTSIPVIVIAYPPIRITSEPTPIYTRPGNTVKLNCMAMGI  
 40 PKADITWELPDKSHLKAGVQARLYGNRFLHPQGSLLTIQHATQRDAGFYKCMKNILGSDS  
 KTTYIHVF

SEQ ID NO: 3

gi|38076521|ref|XM\_143254.3| Mus musculus RIKEN cDNA 6530405F15 gene  
 (6530405F15Rik), mRNA

45 1 atgtgggagc tgggggttcag ggcgagacag aggggtgggat gggcagaagg gtccaggaaa  
 61 aggaagtagc tggaggggag ttgggacaaa agcagcgacc aagggaacat cgcttcagtg  
 121 actgaagcca ggcaaaagga gcgggaagga ttatatgtag cctgggacgc tttcataaac  
 181 actgatgacg tgtttgtgca aagcaagcaa tttgaggaga aacgcctggg acgtcggaaa  
 241 gaaggaaagc gcctccagtt ctctgaagag tcagtcccc agctagtga gactaagcct  
 50 301 actaagcctt ttgctcccgt tggaaagcaa gaacgttcct tcaatcaggt gaaggctctc  
 361 ctcaagaagat ttctgtctc tgcttatgtt acaagaggat tcaaaagcaa gacagaagag  
 421 ctccagatgc agaagagagg cagggaagtc agctgcttgc tgatctccct cactgccatc  
 481 tgcttggtgg tcaccctcgg gacaggggtc tgctctcgcc gatgtgcctg ctatgtgccc  
 541 acagaggtgc actgtacatt tcgggtacctg acctccatcc cagacggcat cccagccaat  
 55 601 gtggaacgag tcaatttagg gtataacagc ctactagat tgacagaaaa tgacttttct  
 661 ggctgagca gactggagtt actcatgctg cacagcaatg gcattcacag agtcagtgac  
 721 aagaccttct cgggcttgca gtccttgacg gtcttaaaaa tgagctataa caaagtccaa

781 ataattgaga aggatacttt gtatggactc aggagcttga cccggttgca cctggatcac  
 841 aacaacattg agtttatcaa ccccgaggcg ttttacggac tcaccttgct ccgcttggtg  
 901 catctagaag gaaaccggct gacaaagctc catccagaca catttgtctc tttgagctat  
 961 cttcagatat ttaaaacctc cttcatttag tacctgtact tgtctgataa cttcctgacc  
 5 1021 tccctcccaa aagaaatggt ctcctctatg ccaaacctag aaagcettta cttgcatgga  
 1081 aaccatgga cctgtgactg ccatttaaag tgggtgtccg agtggatgca gggaaacca  
 1141 gatataataa aatgcaagaa ggaaagaatc cctccagtc ctcagcagtg tcccttttgc  
 1201 atgaacccca ggatctctaa aggcagatct attgctatgg tccatcttg ctcgttcttg  
 1261 tgtacaaagc caaccattga tccatcactg aagtcaaaaa gcctgggtat tcaggaggac  
 10 1321 aatggatctg cctccgtctc acctcaagat ttcatagaac cttttggctc cttgtctttg  
 1381 aacatgacat acctgtctgg aaataaggcc aatgtgatct gtagtatcca aaagccttct  
 1441 aggacattac caattgcatt cactgaagaa aatgactaca tcatgctaaa tatgtcattt  
 1501 tcaacaaatc ttgtgtgcag tgtcaattat aatcacatcc agccagtgtg gcaactcctg  
 1561 gctttgtaca gtgactctcc tctgatatta gaaaggaagc cccagcatac tgagactcca  
 15 1621 ctgctgtctc ccaaatatca acaggtggct cttaggcctg aagacacttt taccaacata  
 1681 gaggtgatt tcaaagcaga tcccttttgg tccaacaag aaaaaatttc cctgcagctc  
 1741 aacagaactg ctaccacact tagcacatta cagatccaat tttccacgga tgctcaaatc  
 1801 actttaccaa aggcagagat gagaccgggt aaacgcaaat ggaccatgat tctgatgat  
 1861 aacaatacca gactggaaca tactgttttg gttggcggca ctattgccct ggactgtccg  
 20 1921 ggcaaagggtg accttcacc tcaattgga tgggttttag ctgatgggag taaagtga  
 1981 gcccttatg ttagtgagga tgggcgaatc ctaatagaca aaaaggggaa gttggaactc  
 2041 cagatggctg atacctttga tgcaggtctt taccattgca taagcaccaa tgatgctgat  
 2101 gcagatattc tcacatacag gataactgtg gtagagccct atgtagaaaa caagcatgaa  
 2161 aatggagctc tgcacacagt aattatgggt gagatactcg atcttccatg cctttccact  
 25 2221 ggtattccag atgcttctat tagctggatg cttccccgga acactgtgtt ctctcagtc  
 2281 tcaagagaca tgcaaatctt taacaatggg accttaagaa tattacaggc tacacaaaa  
 2341 gatcaaggcc attaccgatg ttagcagccc aacctatcag gggctgattt ttccagtttt  
 2401 caagtttcag tccaaatgaa aggtcaaagg acaattgagc atgacaggga catagatgga  
 2461 tctggacttg aagaacccaa gccagtggt ctccttaagc agccaccatc tttgaaactc  
 30 2521 cctgcatcat ctttgacagg gacagaggct ggaaaaaag tctctgggat acataagaaa  
 2581 acaaacata gagacttaac acatcgacgg cgtggggatt ccactctccg gagattcagg  
 2641 gaacacagga ggcagctccc tctctctgct cggagaattg aaaaaaaca ctgggcagca  
 2701 cttttagaaa aagcaaaagaa gaattctgtt ctaagggaagc aagaaaaatc cgcctccggc  
 2761 ccaacgccac tggctattcc acttggtgaa ctcgctgggg aggaaaaaga cgcctccggc  
 35 2821 ctgactctc cagatgaaga attcacgggt ctgaaaaacta aggccttttg tgtcccagaa  
 2881 aggtcaccaa ctgctgactc tagaccagta aatcatggct ttgtgacaag ttcagcttct  
 2941 ggcacagaag tctcctccac cgtgaatccg caaacactac taccacgca ccttccatg  
 3001 tcaaatat ttaattgtgt ggacagtga gctgtgtcaa agagtatgaa ccgacctgta  
 3061 aacagcaaga tagaagatac aacacacaa aaccaatca ttatcttcc atcagtagct  
 40 3121 gaaattcaag attctgctca ggtggggaga acgtcttccc aaagtgcaca ccccgcaaca  
 3181 gggggagcca tggctaccta tggctatacc accatgctta gtagcttcac caacaaagcc  
 3241 aatacagtc tgcagtcagc aaatccaaca gaaagttagt gacctcagat accttaaca  
 3301 gaagtaagta gagttagcag taataactcc ttggctcaca ctactaaaga tccaggttc  
 3361 tccaagcgcc cttcagattc ccacaccact gcccttctt tatttcaaac tcctagaac  
 45 3421 aacagtacag gtaacgtggg aagagagagg acaatttga gcagagggcg agctataagt  
 3481 ccatatagaa ctccagttct cgtcggcat agacacagga ttgtgaggcc agcactcaag  
 3541 ggacctgcta acagaaatat aagtcaagtt tcagccacgg agccccctgg gatgtgccga  
 3601 acctgttctt ccacagaaag gctcaccatg gctacggcag cactgtcagt tacaggctca  
 3661 tcccacacta cctccccaag agctaacaat gttgggatta tttcagaaga gtctaccact  
 50 3721 gtggtcaaga agccatctt actattgaag aacaacaag atgtagatat agagacaata  
 3781 acaaccacta taaactattt cagaagtga agtaccaca tgactcccac tgaagcaagc  
 3841 atgatttctg ctccaacatc catatccctg ggaaaaactc ctatagacac tagtggtcac  
 3901 ctgagcatga ctaggacct ccaagctgga acagatttag tggtagacac accactttcc  
 3961 agtccactta gccaaacctc aatccaaca aaagcaaca gcacaaaact ctcaagaaga  
 55 4021 aaaattccct ggcacccaat ctttgcaaat aaccataaca aagaggggat gctaaagaat  
 4081 ctgcatcagt ttggtttaca aaagaacaca gccacttagc ctctgaaaa agctcctctt  
 4141 ttaccacag atcatgggtc ctcctcacct tctacaacac tcttggaag tctgacgcca  
 4201 gctcagtcag caacaatggc tgccactcgg cgcaatggca ctgaagtga aggtgccaga  
 4261 agtctctctg cagggaaga gcagccctc atcaactcct tctagtgtc tccagcacc  
 60 4321 acaagaaaga gatctagtac attaagctc ctgtcagtg aaacccccac agtgacaact  
 4381 cctcctgtta ttgcatctgc cattatctct gaaacccaag aagtacgac caaaaaagca

4441 aaagacccaaa caaagggggtc tctgaagaac agggaaaggcc caaccatcac ccccaggcag  
 4501 atttctggct atagcacata ctacagttcca acaacaactg atactccctt ggcttccagt  
 4561 cattccccgg gaaaagtcac tggtaggact gtaagtacag ctgctcctca ctacagcagct  
 4621 tctctcctgg gcataactga actgccccag aaatgcactc atacttcggg aaatataaca  
 5 4681 gcttcggaaa caactctggt gagcaaatca caggagagca ccgcaatgaa aagagcctcc  
 4741 gccacaccac cactcctcag cagtggggca ccccgaaatgc ctactccttc ccctcctccc  
 4801 ttcactaagg ttgtgggttac agacagttag gtcccagcag ttttcaagat gatgtcaaat  
 4861 aggatgggtca ccatatatga atcttcaagg cacgatatag atctgcagca accctcagca  
 4921 gaggctagcc ccaatcctga gatcttaact ggatccactg actttccctt ctctagtctg  
 10 4981 ttgacctcca ctccatgccc agcaccaaga gtggataaac cacaggattc tcaatggaag  
 5041 ccttctccct ggccagaaaa caaatctcag ctacaggtcat actcagaaac cattgaaaag  
 5101 ggcaaaaaggc cagaaataag cctgtcaccc cacctcagct ttccagaggc cagactcat  
 5161 gccttgcatg ggaatgcaca gaggcagtga gaaaagagt tctttgataa gaaacctgct  
 5221 caaaacccaa ctcccaaaaca cctgcctat gactctctgc ctaagactat attgaagaaa  
 15 5281 ccaagaataa ttggaggaaa ggctgcaagc tttactgttc caactaatc agatgtcctt  
 5341 ctcccttgcc aggctgttgg agacccaaag cccaccatcc actggaccag agtctcatca  
 5401 ggacgtgaaa tateccgagg gatacagaaa acccggttcc atgtgcttcc caatggcacc  
 5461 ttgtccatcc agaggggtcag cttctcaggac cgtggacagt acctgtgcgc tgcctctaatt  
 5521 ccactagggt tagaccacct tcattgtcact ctgtctgtgg tctcctaccc tcttaggatt  
 20 5581 ctggagagtc atgtcaagga gatcacagct cattccggaa gtactgtgaa acttaaggtg  
 5641 agagtagaag gtatgccaaag acctacaatt tcctggatac tcgcaaacca aacagtgggtc  
 5701 tcggaacacac ccgaggggaa ccggaagggtc tgggtgacac ccgatggaac gttgatcatc  
 5761 cataatttga gtctttatga ccgtggtttt tacaagtgtg tagccaacaa cccatctggc  
 5821 caggattcac tgttggttaa gatacaagtc atcacagtc cccctgttat tatagagcaa  
 25 5881 aagagacaaag ccactcgttgg cgttttaggt gaaagtgtga aactgccttg cactgcaaaag  
 5941 ggaactcccc agcccagtgat tcattgggtc ctctatgatg ggactgaact aaaaccactg  
 6001 cagttgactc attccaggtt tttcttgtat ccaaatggga cccctgtatat aagaaacatc  
 6061 gtttcttcag tcagggggcac ttatgaatgc attgctacca gctcctcggg ctccagagaga  
 6121 agggtagtga ttctaagagt agaagagcaa gagacagttc ccaggataga aactgcctcg  
 30 6181 cagaaatgga cagaggtgaa tttgggtgag aaattactac tgaactgttc agctactggg  
 6241 gatccgaaac ctacaataat ctggaagtta ccatccaagg ttgtcattga ccagtggcac  
 6301 agaattgggca gccgatcca tgtctaccga aatggctcct tgggtatttg atcgggtgaca  
 6361 gaaaaggatg gtgggtgacta ctatgtctgt gcaagaaaaa aaatgggaga tgacctgggtc  
 6421 ctgatgcatg tccgcctaag actgacacct gccaaaattg aacacaagca gcattttaag  
 35 6481 aagcaagtac tccatgggaa agattttcaa gttgactgca aggtctcttg ctccccctgtg  
 6541 cctgaggtat cctggagttt gcctgatgga acagtgggtg acaatgtagc acaagctgat  
 6601 gacagtgggt acaggacca gaggtacacc ctcttccaca acggaacctt gtatttcaac  
 6661 aaagtgtgga tggcagagga aggagattat atttgccttg ctccagaacac cttagggaaa  
 6721 attgaaatga aagttcacct aacagttcta acagccatcc caccgataag gcaaaactac  
 40 6781 agggacaatg taaggatcaa ggctggagac acagtgttcc tggactgtga ggtcgtggg  
 6841 gaacccaagc caaatgtatt ttggtgtctg ccttccaaca atgtcatttc attctccaat  
 6901 gacaggttca tatttcatgc caatggaact ttgtccatca ataaagtga accgcttgac  
 6961 tctgggaagt atgtgtgtgt agctcaaaat cctagtgggg atgacactaa gacatacaaa  
 7021 ctggatattg tctctaggcc tccattaatc aatggtttgt atgcaaacaa aactgttatt  
 45 7081 aaagccacag ccattcagca ctccaaaaaa cacttggact gcagagcaga tgggggtccca  
 7141 ccacccacga tcacatggat tatgccagac aatattttcc tcacggctcc atactacgga  
 7201 ggcagaatca ccgtgcatca aaatggaaac ttggaaatc ggaacataag gctttctgat  
 7261 tctgcggatt tcacctgtgt ggttcggagc gaaggaggag agagtgtgtt ggtggtgcag  
 7321 ttaaaagtac tggaaatgct gagaagacca acattcagaa acccattcaa tgaaaaagta  
 50 7381 gttgcccagg ttggcaagcc ttagcaatg aactgctctg tggatgggaa cccaacacct  
 7441 gaaattatct ggatcttacc tgatggcaca caatttgcta atggggccaca aaattcccca  
 7501 tatctgatgg caagcaatgg ttctctcatt gtttacaagg caactcggaa caagtcagga  
 7561 aagtatcgct gtacagcaag aataaaggtt ggctacatcg agaaactcat cctgttagaa  
 7621 attggacaga agccagtcac tctgacatc gaaccgggga tgataaagag cgccggtggg  
 55 7681 gaactcgctat ccctgcattg tgtgtctgat ggaatcccta agccaaatgt caaatggact  
 7741 acgccagggt gccttgtaac cgacagacct caagtccgtg gaaatatac attgcatgaa  
 7801 aatggcacgt tggcatcaa agaaacaaca gctcatgaca gaggaatta tatctgtaag  
 7861 gctcaaaaca gcgttggcca ggcagttatt agtgttccc tgacgattgt ggcctacctt  
 7921 ccccgaaatta taaactacct tcccaggagc atgctcagga ggacaggaga ggcaatgcag  
 60 7981 ctccactgtg tggccttggg agtccccaa ccacaaatca cctgggagac gccgggatac  
 8041 tccctgtctc caacggcgac agaaagaagg cccacagaa gtgagatgct tcccctacaa

8101 ggtacgctgg tcattcagaa tctacgagcc tcagattctg gcgtctataa atgcagagca  
 8161 cagaacgtgc ttggtgccga ttatgcaaca acttacatcc aggtcctctg acaggaaggg  
 8221 agaaactgaa tggaaacaaa gccaacatct gcagacttta ttttttggaa gaagttaaat  
 8281 caaaggcagc cataggcatg taaatgaatc cgaatacatt tacagtatta aatttacaat  
 5 8341 ggacatgcag tgagacttgt aaataaatgc actgtgaact gattccgagt ttccatggat  
 8401 ttcaaagcaa actcttaact taaggcactt tgattttgcc aacaaataat aaaaaacatt  
 8461 aagagaaaat gatccgctac aaattaacaa atggcgaatg cacctgaatt ttcagtaaaa  
 8521 agacctttct tccgctaaca gttgccagct gccttggtgc tgttttccac caatgttaca  
 8581 aacatcgac acagaatgaa tggagacaac gtgaaagatt aggtttgcag cccgtgttaa  
 10 8641 tctcaatgca caaatatttt gtcactggtt tacaaacatt ttgataaaac ctacagaaaa  
 8701 caagcgcaga actgttctga ttatgattaa tagttttacca tttgttcaca ct

## SEQ ID NO: 4

15 Amino acid sequence of mouse ADLican encoded by the DNA sequence shown in SEQ ID NO: 3.

MWELGFRARQVRVWAEGRKRKRVLEGSWDKSSDQGNIASVTEARQKEREGLYVAWDAFIN  
 TDDVFVQSKQFEEKRLGRRKEGKRLQFSEESVPQLVTKPTKPFAPVGSKERSFNQVKAL  
 LRRFPVSAYVTRGFSKTEELRMQKRGREVSCLLISLTAICLVVTPGSRVCPRRACACYVP  
 20 TEVHCTFRYLTSIPDGI PANVERVNLGYNSLTRLTENDFSGLSRLELLMLHSNGIHRVSD  
 KTFSQLQSLQVLKMSYNKVQII EKDTLYGLRSLTRLHLHDHNNIEFINPEAFYGLTLLRLV  
 HLEGNRLTKLHPDTPVLSYLOIFKTSFIKYLSDNFLTSLPKEMVSSMPNLESLEYLHG  
 NPWTCDCHLKWLSEWMQGNPDII KCKKERIPSSPQQCPLCMNPRISKGRSIAMVPSGSFL  
 CTKPTIDPSLKGKSLGIQEDNGSASVSPQDFIEPFGSLSLNMTDLSGNKANVICS IQKPS  
 25 RLTPIAFTEENDYIMLNMSFSTNLVCSVNYNHIQPVWQLLALYSDSLILERKPQHTETP  
 LLSPKYQQVALRPEDTFTNIEADFKADPFWFQOEKISLQLNRTATTTLSTLQIQFSTDAQI  
 TLPKAEMRPVKRWMTMILMMNTRLEHTVLVGGTIALDCPGKGDPSPHLEWVLADGSKVR  
 APYVSEDGRILIDKKGKLELMADTFDAGLYHCISTNDADADILTYRITVVEPYVENKHE  
 NGALHTVIMGEILDPLCLSTGIPDASISWILPRNTVFSQSSRDMQILNNGTLRLILQATPK  
 30 DQGHYRCVAANPSGADFSFQVSVQMGQRTIEHDDRIDGSGLEPKPSVLLKQPPSLKL  
 PASSLTGTEAGKQVSGIHKKNKHRDLTHRRRGDSTLRRFREHRRQLPLSARRIDPQHWA  
 LLEKAKKNSVLRKQENTTVKPTPLAIPLEVELAGEEKDASGLTPPDEEFTVLKTKAFGVPE  
 RSPTADSRPVNHGFVTSSASGTEVSSTVNPQTLPLTHLPDFKLFNVVDSAAVSKSMNRV  
 TSKIEDTTHQNPIIIFPSVAEQDSAQVGRSTSSQSAHPATGGAMATYGYTTLSSFTNKA  
 35 NTVLQSANPTESYGFQIPLTEVSRVSSNNSLAHTTKDPGFSKRPSSDHTTAPSLFQTPRN  
 NSTGNVGRERTIWSRGRAISPYRTPVLRHRHRIVRPALKGPANRNI SQVSATEPPGMCR  
 TCSSTERLTMTAALSVTGSSHTTLPKANNVGIIESESTTVVKPSLLLNKQDQVDIETI  
 TTTINYFRSESTHMTPEASMISAPTSISLGKTPIDTSGHLSMPRTIQAGTDLVVTPLPS  
 SPLSQPSIPTKATSTKLSRRKIPWHPIFANNHNKEGMLKNLHQFGLQKNTATKPEKAPL  
 40 LPTDHGSSSPSTLLASLTPAQSATMAATRNGTEVQGARSLSAGKEQPFINSFLVLPST  
 TRKRSSTLSFLSVETPTVTPPVIAAI ISETQEVRSKAKDQTKGSLKNRKGPTITPRQ  
 ISGYSTYSVPTTDTPLAFSHSPGKVTGRTVSTAAPHSAASLLGITELPQKCTHTSGNIT  
 ASETLLSKSQESTAMKRASATPPLLSSGAPRMPTSPPPPFTKVVDSEVPVAFKMSN  
 RMVTIYESSRHDIDLQQPSAEASPNPEILTGSTDFPLSSLLTSTPMPAPRVDPKQDSQWK  
 45 PSPWPNKFLRSYSETIEKGKRP EISLSPHLSFPEASTHALHWAQRHAEKSVFDKKPA  
 QNPTSKHLPYDSL PKTILKKPRIIGGKAASFVPTNSDVLPCAEVGD PKPTIHWTRVSS  
 GREISRGIQKTRFHVLPNGTSLIQRVSIQDRGQYLCAASNPLGVDHLHVTL SVVSYPARI  
 LESHVKEITAHSGSTVKLKCRVEGMPRPTISWILANQTVVSETPEGSRKVWVTPDGTLLI  
 50 HNLSLYDRGFYKCVANNPSGQDSLLVKIQVITAPPVIEEQKQAI VGVLGESLKL PCTAK  
 GTPQPSVHVLYDGTTELKPLQLTHSRFFLYPNGTLYIRNIVSSVRGTYECIATSSSGSER  
 RVVILRVEEQETVPRIETASQKWTEVNLGEKLLNCSATGDPKPTIIWKLPSKVVIDQWH  
 RMGSRHVPNGSLVIGSVTEKDG DYLCVARNKMGGDLVLMHVRLRLTPAKIEHKQHF  
 KQVLHGKDFQVDCASGSPVEVSWSLPDGTVNNVAQADDSGYRTKRYTLFHNGLTYFN  
 55 KVGMAEEGDYICSAQNTLGKDEM KVHLTVLTAIPRIRQNYRSNVRIKAGDTAVLDCVETG  
 EPKPNVFWLLPSNNVISFSNDRFI FHANGTSLINKVKPLDSGKYVCVAQNPSGDDTKTYK  
 LDIVSRPPLINGLYANKTVIKATAIQHSHKHLDCRADGVPPPQITWIMPDNIFLTAPYYG  
 GRITVHQNGTLEIRNIRLSDSADFTCVVRSEGGESVLVVLKVLEMLRRPTFRNPFNEKV  
 VAQVGKPVAMNCSVDGNPTPBIIWILPDGTQFANGPQNSPYLMASNGSLIVYKATRNKSG

KYRCTARNKVGYIEKLILLEIGQKPVILTYEFGMIKSAGGESLSLHCVSDGIPKPNVKWT  
 TPGGLVIDRPQVGKYLHENGTLVIKETTAHDRGNVICKAQNSVGQAVISVPVTIVAYP  
 PRIINYLPRSMRLRRTGEAMQLHCVLGVKPKQITWETPGYSLLSTATERRPHRSEMLPLQ  
 GTLVIQNLRASDSGVYKCRAQNVLGADYATTYIQVL

## 5 SEQ ID NO: 5

gi|33355470|gb|AY273816.1| Rattus norvegicus bone specific CMF608 mRNA, complete cds

```

1  cgagagacga  cagaaggtta  cggctgcgag  aagacgacag  aaggggtccag  aaaaaggaaa
61  gtgctggagg  ggagtgggga  caaaagcagc  gaccaagtga  atgtcacttc  agtgactgag
121  gccaggcaaa  acgcgcggga  aggattttgt  gtagcttggg  accctttcat  agacactgat
10  181  gacacgttta  cgcaaaatag  aaattttgag  agaaacgcct  gggccttcgg  aaaggagtga
241  ttgattagta  cttgcaagtt  taggtgactt  taaggagaac  taactaatgt  atactattga
301  gggaggagga  agagcattac  agagtttcca  gcagcagcag  gaaagctttg  gttaatttgg
361  aaatggatga  tagcattaaa  ataacagaag  cgcctccagg  tctctgaagc  ttcagteccc
421  cagctgaaa  ccagaaaaga  ctaagcccac  taagcctttt  gatccctttg  gaagcaaaga
15  481  actttccttc  cctggggtga  agactctctc  cagaagattt  cctgtctctg  cctatgttac
541  aagaggaaac  aaaaccaaga  cagaagagct  caggatgcag  gtgagaggca  ggggaagtcag
601  cggcttggtg  atctccctca  ctgctgtctg  cctgggtggtc  acccctggga  gcaggggcctg
661  tcctcgccgc  tgtgcctgct  atgtgcccac  agaggtgcac  tgtacatttc  ggtacctgac
721  ctccatccca  gatggcatcc  cggccaatgt  ggaacgaata  aatttaggat  ataacagcct
20  781  tactagattg  acagaaaacg  actttgatgg  cctgagcaaa  ctggagttac  tcatgctgca
841  cagtaatggc  attcacagag  tcagtgcaca  gaccttctcg  ggcttgagct  ccttgagagt
901  cttaaaaatg  agctataaca  aagtcctaat  cattcggaag  gatactttct  acggactcgg
961  gagcttggtg  cggttgcacc  tggatcaca  caacattgaa  ttcattcaacc  ctgaggcctt
1021  ttatggactt  acctcgctcc  gcttggtaca  tttagaagga  aaccggctca  caaagctcca
25  1081  tccagacaca  tttgtctcat  taagctatct  ccagatattt  aaaacctctt  tcattaagta
1141  cctgttcttg  tctgataact  tcctgacctc  cctcccaaaa  gaaatggtct  cctacatgcc
1201  aaacctagaa  agcctgtatt  tgcatggaaa  cccatggacc  tgtgactgcc  atttaaagtg
1261  gttgtctgag  tggatgcagg  gaaaccaga  tataataaaa  tgcaagaaag  acagaagctc
1321  ttccagtcct  cagcaatgtc  ccctttgcac  gaacccagg  atctctaaag  gcagaccctt
30  1381  tgctatggta  ccatctggag  ctttctatg  tacaaagcca  accattgatc  catcactgaa
1441  gtcaaagagc  ctggttactc  aggaggacaa  tggatctgcc  tccacctcac  ctcaagattt
1501  catagaaccc  tttggctcct  tgtctttgaa  catgacagac  ctgtctggaa  ataaggccga
1561  catggtctgt  agtatccaaa  agccatcaag  gacatcacca  actgcattca  ctgaagaaaa
1621  tgactacatc  atgctaaatg  cgtcattttc  cacaatatct  gtgtgcagtg  tagattataa
35  1681  tcacatccag  ccagtgtggc  aacttctggc  tttatacagt  gactctcttc  tgatactaga
1741  aaggaagccc  cagcttaccg  agactccttc  actgtcttct  agatataaac  aggtggctct
1801  taggcctgaa  gacattttta  ccagcataga  ggctgatgtc  agagcagacc  cttttgggtt
1861  ccaacaagaa  aaaattgtct  tgcagctgaa  cagaactgcc  accacactta  gcacattaca
1921  gatccagttt  tccactgatg  ctcaaatcgc  tttaccaagg  gcggagatga  gaggcggagag
40  1981  actcaaatgg  accatgatcc  tgatgatgaa  caatcccaaa  ctggaacgca  ctgtcctggg
2041  tggcggcact  attgccctga  gctgtccagg  caaaggcgac  ccttcacctc  acttggaaatg
2101  gcttctagct  gatgggagta  aagtgcagag  cccttacgtt  agcgaggatg  ggcaaatcct
2161  aatagacaaa  aatgggaagt  tggaaactgc  gatggctgac  agctttgatg  caggtcttta
2221  ccactgcata  agcaccaatg  atgcagatgc  ggatgttctc  acatacagga  taactgtggg
45  2281  agagccctat  ggagaaagca  cacatgacag  tggagtccag  cacacagtgg  ttacgggtga
2341  gacgctcgac  cttccatgcc  tttccacggg  tgttccagat  gcttctatta  gctggattct
2401  tccagggaac  actgtgttct  ctcagccatc  aagagacagg  caaattctta  acaatgggac
2461  cttaagaata  ttacagggtta  cgccaaaaga  tcaaggatcat  taccaatgtg  tggctgccaa
2521  cccatcaggg  gccgactttt  ccagttttta  agtttcagtt  caaaagaaag  gccaaaggat
50  2581  ggttgagcat  gacagggagg  caggtggatc  tggacttggg  gaacccaact  ccagtgttct
2641  ccttaagcag  ccagcatctt  tgaactcttc  tgcacagctt  ttgacagggt  cagaggctgg
2701  aaaaacaagtc  tccgggtgtac  ataggaagaa  caaacataga  gacttaatac  atcggcggcg
2761  tggggattcc  acgctccggc  gattcaggga  gcataggagg  cagctccctc  tctctgtctg
2821  gagaattgac  ccgcaacgct  gggcagcact  tctagaaaaa  gccaaaaaga  attctgtgcc
55  2881  aaaaaagcaa  gaaaatacca  cagtaaagcc  agtgccactg  gctgttcccc  tcgtggaact
2941  cactgacgag  gaaaaggatg  cctctggcat  gattcctcca  gatgaagaat  tcatgggtct
3001  gaaaactaag  gcttctgggtg  tcccagggaag  gtcaccaact  gctgactctg  gaccagtaaa

```

	3061	tcattggtttt	atgacgagta	tagcttctgg	cacagaagtc	tcaactgtga	atccacaaac
	3121	actacaatct	gagcaccttc	ctgattttcaa	attattttagt	gtaacaaacg	gtacagctgt
	3181	gacaaagagt	atgaacccat	ccatagcaag	caaaatagaa	gatacaacca	accaaacc
	3241	aatcattatc	tttccatcag	tagctgaaat	tcgagattct	gctcaggcag	gaagagcatc
5	3301	ttcccaaagt	gcacaccctg	taacaggggg	aaacatggct	acctatggcc	ataccaacac
	3361	atatagtagc	tttaccagca	aagccagtac	agtcttgtag	ccaataaatc	caacagaaag
	3421	ttatggacct	cagataccta	ttacaggagt	cagcagacct	agcagtagtg	acatctcttc
	3481	tcacactact	gcagacccta	gcttctccag	tcacccttca	ggttcacaca	ccactgcctc
	3541	gtctttatct	cacattccta	gaaacaacaa	tacaggtaac	ttcccttgtg	ccaggcactt
10	3601	gggaagagag	aggacaattt	ggagcagagg	gagagttaaa	aaccacata	gaacccacgt
	3661	tctccgacgg	catagacaca	ggactgtgag	gccagcaatc	aagggacctg	ctaacaaaaa
	3721	tgtgagccaa	gttccagcca	cagagtaccc	tgggatgtgc	cacacatgtc	cttccgcaga
	3781	ggggctcaca	gtggctactg	cagcactgtc	agttccaagt	tcattccaca	gtgccctccc
	3841	caaaactaat	aatgttgggg	tcatagcaga	agagtctacc	actgtggtca	agaaaccact
15	3901	gttactatct	aaggacaaac	aaaatgtaga	tattgagata	ataacaacca	ctacaaaata
	3961	ttccggaggg	gaaagtaacc	acgtgattcc	tacggaagca	agcatgactt	ctgctccaac
	4021	atctgtatcc	ctggggaaat	ctcctgtaga	caatagtggt	cacctgagca	tgcttgggac
	4081	catccaaact	gggaaagatt	cagtggaaac	aacaccactt	cccagcccc	tcagcacacc
	4141	ctcaatacca	acaagcacia	aattctcaaa	gaggaaaact	cccttgcacc	agatctttgt
20	4201	aaataaccag	aagaaggagg	ggatgttaaa	gaatccatat	caattcgggt	tacaaaagaa
	4261	cccagccgca	aagcttccca	aaatagctcc	tcttttacc	acaggtcaga	gttccccctc
	4321	agattctaca	actctcttga	caagtccgcc	accagctctg	tctacaacaa	tggtgccac
	4381	tcagaacaag	ggcactgaag	tagtatcagg	tgccagaagt	ctctcagcag	ggaagaagca
	4441	gcccttcacc	aactcctctc	cagtgtctcc	tagcaccata	agcaagagat	ctaatacatt
25	4501	aaacttcttg	tcaacggaaa	ccccacag	gacaagtcc	actgtactg	catctgtcat
	4561	tatgtctgaa	acccaacgaa	caagtccaa	agaagcaaaa	gaccaaataa	aggggcctcg
	4621	gaagaacaga	aacaacgcaa	acaccacccc	caggcagggt	tctggctata	gtgcatactc
	4681	agctctaaca	acagctgata	cccccttggc	tttcagtcac	tccccacgac	aagatgatgg
	4741	tggaatgtga	agtgcagttg	cttatcactc	aacaacctct	cttctggcca	taactgaact
30	4801	gtttgagaag	tacaccacga	ctttgggaaa	tacaacagct	ttggaaacaa	cggtgttgag
	4861	caaatacag	gagagtacca	cagtgaagag	agcctcagac	acaccaccac	cactcctcag
	4921	cagtggggcg	ccccacgtgc	ccactccttc	cccactcctt	tttactaagg	gtgtgggtac
	4981	agacagcaaa	gtcacatcag	ctttccagtc	gacgtcaaat	agagtgggtc	ccatatatga
	5041	atcttcaagg	cacaatacag	atctgcagca	accctcagca	gaggctagcc	ccaatcctga
35	5101	gatcataact	ggaaccactg	actctccctc	taatctgttt	ccatccactt	ctgtgccagc
	5161	actaagggtg	gataaaccac	agaattctaa	atggaagccc	tctccctggc	cagaacacaa
	5221	atatcagctc	aagtcatact	ccgaaacccat	tgagaagggc	aaaaggccag	cagtaagcat
	5281	gtccccccac	ctcagccttc	cagaggccag	cactcatgcc	tcacactgga	atacacagaa
	5341	gcattgcagaa	aagagtgttt	ttgataagaa	acctgggtcaa	aacccaactt	ccaaacatct
40	5401	gccttacgtc	tctctaccta	agactctatt	gaaaaagcca	agaataattg	gaggaaaggc
	5461	tgcaagcttt	acagttccag	ctaattcaga	cgtttttctt	ccttgtgagg	ctgtgggaga
	5521	cccactgccc	atcatccact	ggaccagagt	ttcatcagga	cttgaaatat	cccaagggac
	5581	acagaaaagc	cggttccacg	tgcttcccaa	tgccaccttg	tccatccaga	gggtcagtat
	5641	tcaggaccgt	ggacagtacc	tggtgtctgc	atttaatcca	ctgggcgtag	accattttca
45	5701	tgtctctttg	tctgtgtgtt	tttaccgggc	aaggattttg	gacagacatg	tcaaggagat
	5761	cacagttcac	tttggaagta	ctgtggaact	aaagtgcaga	gtggagggtg	tgccgaggcc
	5821	tacggtttcc	tgataacttg	caaaccacaa	gggtgtctca	gaaacggcca	aggggaagcag
	5881	aaagggtctg	gtaacacctg	atggaacatt	gatcatctat	aatctgagtc	tttatgatcg
	5941	tggttttttac	aagtgtgtgg	ccagcaaccc	atctggccag	gattcactgt	tggttaagat
50	6001	acaagtcac	acagctcccc	ctgtcattat	agagcaaaa	aggcaagcca	tcgttggggt
	6061	tttaggtgga	agtttgaaac	tgccctgcac	tgcaaaaagga	actccccagc	ctagtgttca
	6121	ctgggtcctt	tatgatggga	ctgaactaaa	accattgcag	ttgactcatt	ccagattttt
	6181	cttgtatcca	aatggaactc	tgtatataag	aagcatcgct	ccttcagtg	ggggcactta
	6241	tgagtgcatt	gccaccagtc	cctcaggctc	agagagaagg	gtagtgtatc	ttactgtgga
55	6301	agaggggag	acaatcccca	ggatagaaac	tgccctcag	aaatggactg	aggtgaattt
	6361	gggtgagaaa	ttactactga	actgtctcag	tactggggat	ccaaagccta	gaataatctg
	6421	gaggctgcca	tccaaggctg	tcattgcacca	gtggcacaga	atgggcagcc	gaatccacgt
	6481	ctacccaaat	ggatccttgg	tggttgggtc	agtgcaggaa	aaagacgctg	gtgactactt
	6541	atgtgtggca	agaaacaaaa	tggtgagatga	cctagtccctg	atgcatgtcc	gcctgagatt
60	6601	gacacctgcc	aaaattgaac	agaagcagta	ttttaagaag	caagtgtctc	atgggaaaga
	6661	tttccaagtt	gactgcaagg	cctctgggtc	ccctgtgcct	gaggtatcct	ggagtttgcc

6721 tgatgggaca gtgctcaaca atgtagccca agctgatgac agtggctata ggaccaagag  
 6781 gtacaccctt ttccacaatg gaaccttgta tttcaacaac gttgggatgg cagaggaagg  
 6841 agattatata tgctctgccc agaacacctt agggaaagat gagatgaaag tccacctaac  
 5 6901 agttctaaca gccatcccac ggataaggca aagctacaag accaccatga ggctcagggc  
 6961 tggagaaaca gctgtccttg actgcgaggt cactggggaa ccgaagccca atgtattttg  
 7021 gttgctgcct tccaacaatg tcatttcatt ctccaatgac aggttcacat ttcattgcaa  
 7081 tagaactttg tccatccata aagtgaacc acttgactct ggggactatg tgtgcgtagc  
 7141 tcagaatcct agtggggatg acactaagac atacaaactg gacattgtct ctaaactctc  
 10 7201 attaatcaat ggcctgtatg caaacaagac tgtattataa gccacagcca ttcggcactc  
 7261 caaaaaatac tttgactgca gagcagatgg gatcccatct tcccaggtca cgtggattat  
 7321 gccaggcaat attttctctc cagctccata ctttgaagc agagtcaagg tccatccaaa  
 7381 tggaaacctg gagatgagga acatccggct ttctgactct gcggacttca cctgtgtggt  
 7441 tgggagcgag ggaggagaga gttgttgggt agtgagttta gaagtcttag aaatgctgag  
 7501 aagaccaaca ttcagaaacc cattcaacga aaaagtcata gcccagctg gcaagcccg  
 15 7561 agcactgaac tgctctgtgg atgggaacc cccacctgaa attacctgga tcttacctga  
 7621 cggcacacag tttgctaaca gaccacaaa tccccgtat ctgatggcag gcaatggctc  
 7681 tctcatcctt tacaagcaa ctcggaacaa gtcagggag tatcgtctgt cagccaggaa  
 7741 taaggttggc tacatcgaga aactcatcct gttagagatt gggcagaagc cagtcattct  
 7801 gacatacgaa ccagggatgg tgaagagcgt cagtggggaa ccgttatcac tgcattgtgt  
 20 7861 gtctgatggg atccccaagc caaatgtcaa gtggactaca ccgggtggcc atgtaatcga  
 7921 caggcctcaa gtggatggaa aatacatact gcatgaaaat ggcacgctgg tcatcaagc  
 7981 aacaacagct cagaccaag gaaattatat ctgtagggtc caaacagtg ttggccaggc  
 8041 agttatttagc gtgtcagtga tgggtgtggc ctaccctccc cgaatcataa actacctacc  
 8101 caggaacatg ctgaggagga caggggaagc catgcagctc cactgtgtgg ccttgggaat  
 25 8161 cccaagcca aaagtcacct gggagacgcc aagacactcc ctgctctcaa aagcaacagc  
 8221 aagaaaaccc catagaagtg agatgcttca cccacaaggt acgctggtca ttcagaatct  
 8281 ccaaacctcg gattccggag tctataagtg cagagctcag aacctacttg ggactgatta  
 8341 cgcaacaact tacatccagg tactctgaca ggaaggggga gactaaaatt caacagaagt  
 8401 ccacatccac aggggtttatt ttttgaaga agtttaataa aaggcagcca taggcatgta  
 30 8461 aatgagctctg aatacattta cagtattaaa tttacaatgg acatgcgatg agacttgtaa  
 8521 atgaaagcat tgtgaactga aaccgagctc ctgtggatct caaagcaaac tcttaactta  
 8581 aggcactttg attttgccaa caaataataa caaacattaa gagaaaaaaa tgatccacta  
 8641 cgaaataaca aacggctaata gcacctgaat tctcagtaaa aagaccttcc tctcgctaac  
 8701 agttgccagc tgcctcgtgt ctgtttccta ccaatgtcac aaacatcgca cacagggtga  
 35 8761 atggagctca cgggaaagat taagtttgcg gtctgtgtaa atctcaatgt acaaatattc  
 8821 tgtcnctggg ttataaacat tttgataaaa ccgaaaaaaa aaaaaaaaaa aaaaaaaaaa  
 8881 aaa

SEQ ID NO: 6

40 Amino acid sequence of rat ADL1CAN encoded by the DNA sequence shown in SEQ ID NO: 5.

MQVRGREGVSGLLISLTAVCLVVTGPSRACPRRCACYVPTEVHCTFRYLTSIPDGIPANVE  
 RINLGYNLSLRLTENDFDGLSKLELLMLHSNGIHRVSDKTFSGLSQLQVLKMSYNKVQII  
 RKDTFYGLGSLVRLHLDHNNIEFINPEAFYGLTSLRLVHLEGNRLTKLHPDTFVSLSYLQ  
 45 IFKTSFIKYLFLSDNFLTSLPKEMVSYMPNLESYLHGNPWTCDCHLKWLSEWMQGNPDI  
 IKCKKDRSSSPQCPLCMNPRISKGRPFAMVPSGAPLCTKPTIDPSLKSLSLVTQEDNG  
 SASTSPQDFIEPFGSLNMTDLGKADMVCSIQKPSRTSPTAFTEENDYIMLNASFST  
 NLVCSVDYNHIQPVWQLLALYSDSLILERKQPOLTETPSLSSRYKQVALRPEDIFTSIEA  
 DVRADPFWFQKEKIVLQLNRTATTLSTLQIQFSTDAQIALPRAEMRAERLKWTMILMMNN  
 PKLERTVLVGTTIALSCPGKGDPSPHLEWLLADGSKVRAPYVSEDGRILIDKNGKLELQM  
 50 ADSFDAGLYHCISTNDADADVLTYRITVVEPYGESTHDSGVQHTVVTGETLDLPCLSTGV  
 PDASISWILPGNTVFSQPSRDRQILNNGTLRILQVTPKDQGHYQCVANPSGADFSSFKV  
 SVQKKGQRMVEHDREAGGSGLGEPNSSVSLKQPASLKSASALTGSEAGKQVSGVHRKNK  
 HRDLIHRRGDSLRRFRHRRQLPLSARRIDPQRWAALLEKAKKNSVPPKQENTTVKPV  
 PLAVPLVELTDEEKDASGMIPDEEFMVLKTKASGVPGRSPTADSGPVNHGFMSTIASGT  
 55 EVSTVNPQTLQSEHLPDFKLFVSTNGTAVTKSMNPSIASKIEDTTNQNPPIIFPSVAEIR  
 DSAQAGRASSQSAHPVTGGNMATYGHTNTYSSFTSKASTVLQPINPTESYGPQIPITGVS  
 RPSSSDISSHTADPSFSSHPSGSHTTASSLFHIPRNNNTGNFPLSRHLGRERTIWSRGR



VKNPHRTPVLRHRHRTVRPAIKGPANKVNSQVPATEYPGMCHTCPSAEGTLVATAALSV  
 PSSSHSALPKTNNVGVIAEESTTVVKKPLLLPKDKQNVDBIITTTTKYSGGESNHVIPT  
 EASMTSAPTSVSLGKSPVDNSGHLSPGTIQTGKDSVETPLPSPLSTPSIPTSTKFSKR  
 5 KTPLHQIFVNNQKKEGMLKNPYQFGLQKNPAAKLPKIAPLLPTGQSSPSDSTLLTSPPP  
 ALSTTMAATQNKGTVEVSGARSLASAGKKQFTNSSPVLPTISKRSNTLNFLSTETPTVT  
 SPTATASVIMSETQRTSRKEAKDQIKGPRKNRNNANTTPRQVSGYSAYSALTADTPLAF  
 SHSPRQDDGGNVSAVAYHSTTSLLAITELFEKYTQTLGNTTALETTLLSKSQESTTVKRA  
 SDTPPPLLSSGAPPVPTSPPPFTKGVTDSKVTSAFQMTSNRVVTIYESSRHNTDLQOP  
 10 SAEASPNEIITGTTDPSNLPSTSVPALRVDKPNQSKWKPSWPPEHKYQLKSYSETIE  
 KGKRPVAVSMSPHLSLPEASTHASHWNTQKHAEKSVFDKPKGQNPSTSKHLPYVSLPKTLLK  
 KPRIIGGKAASFVTPANSDFVLPCEAVGDPLPIIHWTRVSSGLEISQGTQKSRFHVLPNG  
 TLSIQRVSIQDRGQYLCSAFNPLGVDHFVLSLVFYPARILDRHVKEITVHFGSTVELK  
 CRVEGMPRPTVSWILANQTVVSETAKGSRKVWVTPDGTLLIYNLSLYDRGFYKCVASNP  
 15 GQDSLLVKIQVITAPVPIEBQKRQAIQVGLGGSLKLPCTAKGTPQPSVHWVLYDGTCLKP  
 LQLTHSRFFLYPNGTLYIRSIAPSVRGTYECIATSSSGSERRVILTVEEGETIPRIETA  
 SQKWTEVNLGKELLLNCSATGDPKPRIIWRLPKAVIDQWHRMGSRIVHPNGSLVVGVS  
 TEKDAGDYLCVARNKMDDDLVLMHVRLRLTPAKIEQKQYFKQVLHGKDFQVDCASGSP  
 VPEVSWSLPDGTVLNNAQADDGSGYRTKRYTLFHNGLTYFNNVGMAGEGDIYICSAQNTLG  
 KDEMKVHLTVLTAIPRIQSYKTTMLRAGETAVIDCEVTGEPKPNVFWLLPSNNVISPS  
 20 NDRFTFHANRTLSIHKVKPLDSGDYVCVAQNPSGDDTKTYKLDIVSKPPLINGLYANKTV  
 IKATAIRHSKKYFDCRADGIPSSQVTWIMPGNIFLPAPYFGRVTVHPNGTLEMRNIRLS  
 DSADFTCVVRSEGGESVLVVQLEVLEMLRRPTFRNPFNEKVIAQAGKPVALNCSVDGNPP  
 PEITWILPDGTQFANRPHNSPYLMAGNGLILYKATRNKSGKYRCAARNKVGYIEKLILL  
 EIGQKPVILTYEPCMVKSVSSEPLSLHCVSDGIPKPNVKTTPGGHVIDRPQVDGKYILH  
 25 ENGLTVIKATTAHQDQNYICRAQNSVGQAVISVSMVMVAYPPRIINYLPRNMLRRTGEAM  
 QLHCVALGIPKPKVTWETPRHSLLSKATARKPHRSEMLHPQGTLVIQNLQTSDSGVYKCR  
 AQNLLGTDYATTYIQVL

SEQ ID NO: 7

gi|4502040|ref|NM\_000693.1| Homo sapiens aldehyde dehydrogenase 1 family, member A3  
 30 (ALDH1A3), mRNA

1 agccgggtgcg ccgcagacta gggcgccctcg ggccagggag cgcggaggag ccatggccac  
 61 cgctaacggg gccgtggaaa acgggcagcc ggacgggaag ccgcccggccc tgccgcgccc  
 121 catccgcaac ctggaggtca agttcaccaa gatatttatc aacaatgaat ggcacgaatc  
 181 caagagtggg aaaaagtgtg ctacatgtaa cccttcaact cgggagcaaa tatgtgaagt  
 35 241 ggaagaagga gataagcccg acgtggacaa ggctgtggag gctgcacagg ttgccttcca  
 301 gaggggctcg ccatggcgcc ggctggatgc cctgagtcgt gggcggtgc tgcaccagct  
 361 ggctgacctg gtggagaggg accgcgccac cttggccgcc ctggagacga tggatacagg  
 421 gaagccattt cttcatgctt ttttcacga cctggagggc tgtattagaa ccctcagata  
 481 ctttgacagg tgggcagaca aaatccaggg caagaccatc cccacagatg acaacgtcgt  
 40 541 atgcttcacc aggcattgag ccattgtgtg ctgtggggcc atcactccat ggaacttccc  
 601 cctgctgatg ctgggtgtga agctggcacc cgccctctgc tgtgggaaca ccatggtcct  
 661 gaagcctcg gaggcagacac ctctcaccgc cttttatctc ggctctctga tcaaagaggc  
 721 cgggttcctt ccaggagtgg tgaacattgt gccaggatc gggccacag tgggagcagc  
 781 aatttcttct caccctcaga tcaacaagat cgccttcacc ggctccacag aggttgga  
 45 841 actggttaaa gaagctgctg cccggagcaa tctgaagcgg gtgacgctgg agctgggggg  
 901 gaagaacccc tgcacgtgtg gtgcggacgc tgacttgga tggcagtg agtggtgcca  
 961 tcaggcagtg ttcttcaacc aaggcagga ttgcacgga gcctccaggg tgttctgga  
 1021 ggagcaggtc tactctgagt ttgtcaggcg gagcgtggag tatgccaa aacggccgt  
 1081 gggagacccc ttcatgtgca aaacagaaca ggggcctcag attgatcaa agcagttcga  
 50 1141 caaaatctta gagctgatcg agagtgggaa gaaggaagg gccaagctgg aatgcggggg  
 1201 ctacagccatg gaagacaagg ggctcttcac caaaccact gtcttctcag aagtcacaga  
 1261 caacatgagg attgccaaag aggagatttt cgggccagt caaccaatac tgaagttcaa  
 1321 aagtatcgaa gaagtataa aaagagcgaa tagcaccgac tatggactca cagcagccgt  
 1381 gttcacaaaa aatctcgaca aagcctgaa gttggcttct gccttagagt ctggaacggg  
 55 1441 ctggatcaac tgctacaacg ccctctatgc acaggctcca tttgggtggc ttaaaatgtc  
 1501 aggaatggc agagaactag gtgaatacgc tttggccgaa tacacagaag tgaactgt  
 1561 caccatcaaa cttggcgaca agaaccctg aaggaaaggc ggggctcctt cctcaaacat

```

1621 cggacggcgg aatgtggcag atgaaatgtg ctggaggaaa aaaatgacat ttctgacctt
1681 cccggggacac attcttcttg aggcctttaca tctactggag ttgaatgatt gctgttttcc
1741 tctcactctc ctgtttatcc accagactgg ggatgcctat aggttgctctg tgaatcgca
1801 gtccctgcctg gggagggagc tgttgggccat ttctgtgttt ccctttaaac cagatcctgg
5 1861 agacagttag atactcaggg cgttggttaac agggagtggg atttgaagtg tccagcagtt
1921 gcttgaatg ctttgccgaa tctgactcca gtaagaatgt gggaaaacc cctgtgtgtt
1981 ctgcaagcag ggctcttgca ccagcggctt cctcagggtg gacctgctta cagagcaagc
2041 cagcctctt tccgaggtga aggtgggacc attccttggg aaaggattca cagtaagggtt
2101 ttttggtttt tgttttttgt tttcttgttt ttaaaaaaag gatttcacag tgagaaagtt
10 2161 ttggttagtg cataccgtgg aagggcgcca gggctcttgt ggattgcatg ttgacattga
2221 ccgtgagatt cggcttcaaa ccaatactgc ctttggaata tgacagaatc aatagcccag
2281 agagcttagt caaagacgat atcacggctt accttaacca aggcactttc ttaagcagaa
2341 aatattgttg aggttacctt tgcctgctaaa gatccaatct tctaaccgca caacagcata
2401 gcaaatccta ggataattca cctcctcatt tgacaaatca gagctgtaat tcaatttaac
15 2461 aaattacgca tttctatcac gtctactaac agcttatgat aagtctgtgt agtcttcctt
2521 ttctccagtt ctgttaccga atttagatta gtaaagcgta cacaactgga aagactgctg
2581 taataacaca gccttggttat ttttaagtcc tattttgata ttaatttctg attagttagt
2641 aaataacacc tggattctat ggaggacctc ggtcttcac ccaagtggcct gagtatttca
2701 ctggcagggtt gtgaattttt cttttcctct ttgggaatcc aaatgatgat gtgcaatttc
20 2761 atgttttaac ttgggaaact gaaagtgttc ccatatagct tcaaaaacaa aaacaaatgt
2821 gttatccgac ggatactttt atggttacta actagtactt tcctaattgg gaaagttagt
2881 cttaagtgtt caaattaagt tggggagggc aataataaaa tgagggcccg taacagaacc
2941 agtgtgtgta taacgaaaac catgtataaa atgggcctat cacccttgct agagatataa
3001 attaccacat ttggcttccc ttcacagct aacacttatc acttatacta ccaataactt
25 3061 gttaaatcag gatttggtct catacactga attttcagta ttttatctca agtagatata
3121 gacactaacc ttgatagtga tacgttagag ggttcctatt cttccattgt acgataatgt
3181 ctttttaacg aaatgctaca ttatttataa ttggtagagt tattgtatct ttttatagtt
3241 gtaagtacac agaggtggtg tatttaaac tctgtaatat actgtattta gaaatggaaa
3301 tatatatagt gttaggttcc acttctttta aggtttaccc ctgtggtgtg gtttaaaaat
30 3361 ctataggcct gggaattccg atcctagctg cagatcgcat ccacaaatgc gagaatgata
3421 aaataaaatt ggatatttga ga

```

SEQ ID NO: 8

Amino acid sequence of human ALDH1A3 encoded by the DNA sequence shown in SEQ ID NO: 7.

```

35 MATANGAVENGQPDGKPPALPRPIRNLEVKFTKIFINNEWHESKSGKKFATCNPSTREQI
CEVEEGDKPDVDKAVEAAQVAFQRGSPWRRDLALSRGRLHQLADLVERDRATLALETM
DTGKPFLLHAFFIDLEGCIRTLRYFAGWADKIQQKTIPTDDNVVCFTRHEPIGVCGAITPW
NFPLMLLVKLAPALCCGNTMVLKPAEQTPLTALYLGSLIKEAGFPVGVNIVPGFGPTV
GAAISSHPQINKIAFTGSTVEGKLVKEAASRNLKRVTLLELGGKNPCIVCADADLDLAVE
40 CAHQGVFFNQQCCTAASRVFVEEQVYSEFVRSVEYAKRPVGDVDFDKTEQGPQIDQK
QFDKILELIESGKKEGAKLECGGSAMEDKGLFIKPTVFSEVTDNMRIAKEEIFGPVQPIL
KFKSIEEVIKRANSTDYGLTAAVFTKNLDKALKLASALESGTVWINCYNALYAQAPFGGF
KMSGNGRELGEYALAEYTEVKTVTIKLGDKNP

```

SEQ ID NO: 9

45 gi|9295527|gb|AF280404.1|AF280404 Mus musculus retinaldehyde dehydrogenase 3 (Raldh3) mRNA, complete cds

```

1 gagagtgcga accagttatg gctaccacca acggggctgt ggaaaacgga cagccggatg
61 ggaaaaccgc tgccttgccg cgccccatcc gcaacttgga ggtcaagttc accaagatat
121 ttatcaacaa cgactggcac gaatccaaga gtggaagaaa gtttgccaca tataaccctt
50 181 caacactaga gaaaatatgt gaggtggaag aaggagataa gcccgatgtg gacaaggctg
241 tggaggccgc tcaagctgcc ttccagcggg gatccccgtg gcgccggctg gatgcactga
301 gcagaggcca gttgctgcat cagctggctg accttgtaga aagggaccgt gcgatcctgg

```

```

361 ctactctgga gaccatggac accggcaagc cattccttca tgcctttttc gtcgacctgg
421 aaggctgtat taagaccttc agatattttg ccgggtgggc agacaaaatc cagggcagga
481 ccatccccac agatgacaac gttgtgtgct tcaccaggca tgagcccatc ggggtgtgtg
5 541 gggccattac accatggaac ttccccctgc tgatgctggc ctggaaactg gctcctgccc
601 tgtgctgtgg gaacaccgtg gtcctgaagc cagctgagca gacctctctc acggctctgt
661 acctagcctc tctcatcaaa gaggtcgggt tccctccggg tgtgggtaac attgtaccag
721 gctttggggc cactgtggga gcagcaattt cctcccatcc gcagatcaac aagatagcct
781 tcaccggctc cacagaggtt ggaaagctgg tcagagaagc cgctcccgag agcaacctga
841 agagggtcac actggagcta ggaggcaaga atccgtgcat cgtgtgtgca gatgctgact
10 901 tggacttggc cgtcgagtgt gctcaccagg gagtgttctt caaccaaggc cagtgtgtga
961 cagcggcctc caggggtgtt gtggaagagc aggtctacgg ggagtttgtg aggaggagtg
1021 tggagttcgc caagaagagg ccggttggag accccttcga tgccaaaacg gagcaggggc
1081 ctcagatcga ccaaaagcag tttgacaaaa tctcagagct gattgagagt gggaagaagg
1141 aaggggcca gctagaatgt ggggggtcag ccatggagga cagagggtg ttcataaac
15 1201 ccacggtctt ctcagatgtt acggacaaca tgaggattgc caaaggagg attttcggac
1261 cagtgcagcc gatcctgaag ttcaaaaacc tggaggaggt gatcaagaga gcgaatagca
1321 ccgactatgg actcacagca gcagtgttca ccaaaaacct ggacaaagca ctgaagctgg
1381 ctgctcgctc cgagtcgggg acagtcctgga tcaactgcta caatgcattt tatgcacagg
1441 ctccatttgg tggcttcaaa atgtctggga atggcagaga actaggagaa tatgctctgg
20 1501 ctgaatatat agaagtgaac accgtcacca tcaaaactga ggagaagaac ccctgaggaa
1561 cag

```

SEQ ID NO: 10

Amino acid sequence of mouse ALDH1A3 encoded by the DNA sequence shown in SEQ ID NO: 9.

```

25 MATTNGAVENGQPDGKPPALPRPIRNLEVKFTKIFINNDWHESKSGRK FATYNPSTLEKI
CEVEEGDKPDVDKAVEAAQAAFQRGSPWRRDLALSRGQLLHQLADLVERDRAILATLETM
DTGKPFLLHAFVDLEGGIKTFRYFAGWADKIQRTIPTDDNVVCFTRHEP IGVCGAITPW
NFPLLMLAWKLAPALCCGNTVVLKPABQTPLTALYLASLIKEVGFPPGVVNIVPGFGPTV
30 GAAISSHPQINKIAFTGSTEVGKLVREAAASRNLKRVTLLEGGKNPCIVCADADLDLAVE
CAHQGVFFNQGCCTAASRVFVEEQVYGEFVRRSVEFAKKRPVGD PFDKTEQGPQIDQK
QFDKILELIESGKKEGAKLECGGSAMEDRGLFIKPTVFSVDVTDNMRIAKEEIFGPVQPIL
KFKNLEEVIKRANSTDYGLTAAVFTKNLDKALKLAAALESGTVWINCYNFYAQAPFGGF
KMSGNGRELGEYALAEYTEVKTVTIKLEEKNP

```

SEQ ID NO: 11

35 gi|23463282|ref|NM\_153300.1| Rattus norvegicus aldehyde dehydrogenase family 1, subfamily A3 (Aldh1a3), mRNA

```

1 atggccaccg ctaacggggc cgtggaaaac ggacagcctg atgggaaacc gcctgccttg
61 ccgcgcccc a tccgcaactt ggaggtcaag ttcactaaga tatttatcaa caatgactgg
121 cacgaaccca agagtggaag aaagtttgcc acatataacc cttcaactct agagaaaata
40 181 tgtgaggtgg aagaaggcga taagccgat gtggacaagg ccgtggaggc cgcccaagct
241 gccttccaga ggggctcccc gtggcgagg ctggatgccc tgagtcgtgg ccagttgttg
301 caccagttgg ctgatcttat agaaagggc cgcgctatcc tggcaactct agagaccatg
361 gacaccggca agccgttcc tcatgcctt ttcgtcgacc tggagggtg tattaagacc
421 ttcagatatt ttgccgggtg ggcagacaaa atccaggggc ggaccatccc cacagatgat
45 481 aatgtcatgt gcttcaccag gcatgagccc attgggggtg gtggggccat tacaccatgg
541 aacttcccc t gttgatgct ggcctggaaa ctggctcctg ccctgtgtg cgggaacacc
601 gtggtcctga agccagctga gcagacgct ctcactgccc tgtacctcgc ttctctcatc
661 aaagaggctg ggttccctcc aggtgtggtg aacattgtgc caggctttgg gccacagtg
721 ggagcagcaa tctcctctca tccacagatc aacaagatag ccttcaccgg ctccacagag
50 781 gttggaaagc tgggtaaaga agctgcctcc aggagcaacc tgaagcgggt cacactggag
841 ctgggaggca ggaacccgtg catcgtgtgt gcggacgctg acctggactt ggctgtggag
901 tgtgctcacc agggagtgtt cttcaaccaa ggccagtgtg gcacagcggc ctccagggtg

```

961 tttgtggaag agcaggtcta cggggagttt gtgaggagga gtgtggagtt cgccaagaag  
 1021 agggccagttg gagacccctt cgatgccaaa acggagcagg ggcctcagat cgaccaaaag  
 1081 cagtttgaca aaatcctcga gctgatcgag agtgggaaga aggaaggggc caagctagaa  
 1141 tgtggggggt cagccatgga ggacagaggg ctgttcatca aaccaccgt cttctcagac  
 5 1201 gttacggaca acatgaggat tgccaaagag gagatttttg gaccagtgc accaatactg  
 1261 aagttcaaaa acctggagga agtaataaaa agagcgaaca gcactgacta tgggctcacg  
 1321 gcggccgtgt tcacaaaaaa cctcgacaaa gccctgaagc tggcttccgc gctcgagtgc  
 1381 gggacagtct gggtaactg ctacaacgca ttctatgcac aggtccatt tgggtgcttc  
 1441 aaaatgtctg gaaatggcag agaactaggt gaatacgctt tgggtgaata cacagaagtg  
 10 1501 aaaactgtca ccatcaaact tgatgagaag aaccctga

SEQ ID NO: 12

Amino acid sequence of rat ALDH1A3 encoded by the DNA sequence shown in SEQ ID NO: 11.

15 MATANGAVENGQPDGKPPALPRPIRNLEVKFTKIFINNDWHEPKSGRKPFATYNPSTLEKI  
 CEVEEGDKPDVDKAVEAAQAFQRGSPWRRILDALSRGQLLHQLADLIERDRAILATLETM  
 DTGKPLHAFFVDLEGCIKTRYFAGWADKIQGRTIPTDDNVMCFTRHEPIGVCGAITPW  
 NFPLMLAWKLAPALCCGNTVVLKPAEQTLTALYLASLIKEVGFPPGVVNI VPGFGPTV  
 20 GAAISSHPQINKIAFTGSTEVGKLVKEAASRSNLKRVTLGLGRNPCIVCADADLDLAVE  
 CAHQGVFFNQQCCTAASRVFVEEQVYGEFVRRSVEFAKKRPVGD PFDKTEQGPQIDQK  
 QPDKILELIESGKKEGAKLECGGSAMEDRGLFIKPTVFSVDTDNMRIAKEEIFGPVQPIL  
 KFKNLEEVIKRANSTDYGLTAAVFTKNLDKALKLASALESGTVWVNCYNFYAQAPFGGF  
 KMSGNGRELGEYALAEYTEVKTVTIKLDEKNP

SEQ ID NO: 13

25 gi|21687097|ref|NM\_145259.1| Homo sapiens activin A receptor, type IC (ACVR1C),  
 mRNA

1 ggtcaccgcc cggctgcggg gccagtgcca ggagcgccac gcaccgccag ccgcaggggg  
 61 cgtgggatgg gggcgggcgg ggaggggggc gcccacactg actagagcca accgcgcact  
 121 tcaaaagggg gtccggtgcc cgctcccctc ccgcggcccg ggaacttcaa agcggggcgt  
 181 gctgccccgg ctgcctcgct ctgctctggg gcctcgagc cccggcgcgg ccgcctgggt  
 30 241 gcgatgacct gggcgctctg ctacgcgctc cgccaggctc tctgtctgct cgcagcggcc  
 301 cccgagctct cgccaggact gaagtgtgta tgtcttttgt gtgattcttc aaactttacc  
 361 tgccaaacag aaggagcatg ttgggcatca gtcattgtaa ccaatggaaa agagcagggtg  
 421 atcaaatcct gtgtctccct tccagaactg aatgctcaag tcttctgtca tagttccaac  
 481 aatgttacca aaaccgaatg ctgcttcaca gatttttgca acaacataac actgcacctt  
 35 541 ccaacagcat caccaaatgc cccaaaactt ggacccatgg agctggccat cattattact  
 601 gtgcctggtt gcctcctgct catagctgag atgctgacag tatgggcatg ccagggtcga  
 661 cagtgtcctt acaggaagaa aaagagacca aatgtggagg aaccactctc tgagtgaat  
 721 ctggtaaatg ctggaaaaac tctgaaagat ctgatttatg atgtgacgac ctctggatct  
 781 ggctctggtc tacctctggt gggttcaagg acaattgcaa ggacgattgt gcttcaggaa  
 40 841 atagtaggaa aaggtagatt tggtaggtg tggcatggaa gatggtgtgg ggaagatgtg  
 901 gctgtgaaaa tattctcctc cagagatgaa agatatttgt ttcgtgaggc agaaatttac  
 961 cagacgggtc tgctgcgaca tgaaaacatc cttggtttca ttgctgctga caacaaagat  
 1021 aatggaactt ggactcaact ttggctggta tctgaatc atgaacaggg ctccttatat  
 1081 gactatttga atagaaatat agtgaccatg gctggaatga tcaagctggc gctctcaatt  
 45 1141 gctagtgggc tggcacacct tcatatggag attgttggtg cacaaggtaa acctgtatt  
 1201 gctatcgag acataaaatc aaagaatc ttagtaaaaa agtgtgaaac ttgtgccata  
 1261 gcggacttag ggttggctgt gaagcatgat tcaatactga acaccatcga catacctcag  
 1321 aatcctaaag tgggaaccaa gaggtatag gtcctgaaa tgcttgatga tacaatgaat  
 1381 gtgaatatct ttgagtcctt caaacgagct gacatctatt ctgttggtct gggttactgg  
 50 1441 gaaatagccc ggaggtgttc agtcggagga attgttgagg agtaccaatt gccttattat  
 1501 gacatgggtg cttcagatcc ctcatagag gaaatgagaa aggttggttg tgaccagaag  
 1561 tttcgaccaa gtatcccaaa ccagtggcaa agttgtgaag cactccgagt catggggaga

```

1621 ataatgcgtg agtggttgga tgccaacgga gcggcccgcc taactgctct tcgtattaag
1681 aagactatat ctcaactttg tgtcaaagaa gactgcaaag cctaattgat ataattatgt
1741 taaaaagaaa tctctcatag ctttcttttc cattttcccc tttatgtgaa tgtttttgcc
1801 attttttttt tgttctacct caaagataag acagtacagt atttaagtgc ccataaggca
5 1861 gcatgaaaag ataactctaa agttaagcat gggcaggagt tgacttcac ccaatctctat
1921 gttatgttta attttatttt gaaagcaaca cctcaactca tctttttatt taataaggaa
1981 gaaatatatt acaaaagtat aaaataagct ctataaaaat gttatagtca ttaagttttt
2041 attttacttg aaccaagagc acatgaatga acaggaaaag atgtaaaaac atttttttct
2101 gagatgaaaa catattaatt aaacatgcaa attagagcat gctatcttta ggtgatgcaa
10 2161 tctatgtttc ccccttttta agtttagcagg acttttttaa aataaatatt gctctaaact
2221 ttaatatatc gaacgtgaga gtggagctgc ttagtggaag atgtaagtga ggtgggtgct
2281 ccatgtgctt ggtctccctt tctgtgttct tctgttctt cataatccac tactgcagca
2341 gtccctgaac cactaaactt gttcctttca tttacaaaag agatacctga catcctgaga
2401 cactgagaaa tgtcctgaag tcacacagct aatggcagaa ctggcactag gtccaaatct
15 2461 tgtgataatg aacaccgtaa ggtagctag ctctctactt tcccttgaat agtgcttttc
2521 tccctatgta atatctttta ttatgatatt tgtggtttag aaggcatatt gagttatttt
2581 gcagaatcat aatggaccg caaaaaatct cagaaccata tctgttgaca ttttttctca
2641 tagaaatatt atggttacc catttgtaa tgagcattaa tgttttctga acacttccaa
2701 agattaatca aacataaata ttcattgtct gaaaatgtct ttaagatata attcagaggt
20 2761 cctattttcc tttgtacata cacacttaga aagaaaagac agaaaaggaa gaggaaggaa
2821 ggaaatattt tgagaatata ttgagaagaa ttaagaaaac tcttcaatga agtggttaaca
2881 accaaaccct acagacggtt tcagaaacag caaatagata tctctctacc ctttcacagt
2941 gagtgagtga gtacagaaga atgctcatga tagttttgcc ttcattctac tttctgtgga
3001 cacagagtaa tgaatattta atgggacatt aaatatgccc ttcaaactta taattttact
25 3061 ttggtaaacg agatttaaca tgatgtcttt tatgctccta aaacatcttt tttcaaacte
3121 cattccttag aacattcttc tactgagatg atccaagacc aaaagtgttc tttggtactt
3181 gcttataaag tgatagtaca tgtagcata taatgtattt tgaagagtga agtaaagtct
3241 attgataaca gaaaaaaaa aaaaaaa

```

SEQ ID NO: 14

30 Amino acid sequence of human ALK7 encoded by the DNA sequence shown in SEQ ID NO: 13.

```

MTRALCSALRQALLLLAAAEELSPGLKCVLLCDSSNFTCQTEGACWASVMLTNGKEQVI
KSCVSLPELNAQVFCSSNNVTKECCFTDFCNNTLHLPTASPNAPKLGPMELAIITV
PVCLLSIAAMLTWACQGRQCSYRKKRPNVEEPLSECNLVNAGKTLKDLIYDVTASGSG
35 SGLPLLQRTIARTIVLQEI VQKGRFGEVWHGRWCGEDVAVKIFSSRDERYWFREAEIYQ
TVMLRHENILGFIAADNKDNGTWTQLWLVS EYHEQGSLYDYLNRNIVTMAGMIKLALSIA
SGLAHLHMEIVGTQKPAIAHRDIKSKNILVKCETCAIADLGLAVKHDSILNTIDIPQN
PKVGTTRYMAPEMLD DDTMNVNIFESPKRADIYSVGLVYWEIARRCSVGGIVEEYQLPYD
MVPSPDPSIEEMRKVVCDQKFRPSIPNQWQSCEALRV MGRIMRECWYANGAARLTALRIKK
40 TISQLCVKEDCKA

```

SEQ ID NO: 15

gi|38074652|ref|XM\_194020.3| Mus musculus activin A receptor, type IC (Acvr1c), mRNA

```

1 gccccgggaa cttcaaagcg gcccgcgctg cgggctgcgc tctgggaccc cgaagccttg
61 caccgcccgc ggggtggccat gaccccgagc gcgggctccg cactgagcct ggccctcttg
45 121 ctgggtggcc tggccgcca ccttgcgcca ggactgaagt gtgtgtgtct tttgtgtgat
181 tcttcaaact tcacctgcca aacggaagga gcatgtggg cctccgtcat gctaaccaac
241 gggaaaagagc aggtgatcaa atcgtgtgtg tccctcccg aactaaatgc tcaggctctc
301 tgtcacagtt ccaacaacgt gacaaaacc gaatgttgc tcaagactt ctgcaacaat
361 atcacactgc acctccac agcatcgcca aatgccctc gacttggtcc cacagagctg
50 421 acggttggtt tcacctgccc ggtttgcctc ctgtccatag ccgccatgct aacaatatgg
481 gcgtgccagg accgccagtg cacatacagg aagactaaga gacacaacgt ggaggaagcg
541 ctggctgagt acagccttgt gaatgcagga aaaacactca aggatctgat ctacgatgct
601 actgcctctg gtcaggctc tggctgcct ctcttggttc aaagaacaat cgcaaggaca

```

661 attgtacttc aagaaatcgt aggaaaaggt cggtttgggg aagtgtggca cggaagatgg  
 721 tgtggagaag atgtggctgt gaaaatattc tcctccaggg atgagagatc ttggttccgt  
 781 gaggcggaaa tttatcagac ggtgatgctg agacacgaga acatcctcgg tttcatcgca  
 841 gctgacaaca aagataatgg aacttgga caactctggc ttgtgtcaga gtatcacgag  
 5 901 cagggtcctt tgtatgacta tttgaacaga aacatagtga ctgtggctgg aatgggtcaag  
 961 ctgggcgttt ccatagcgag tgggtctggct cacctgcaca tggagatcgt gggactcaa  
 1021 ggtaagcctg ctattgctca ccgagatata aagtcaaaga atatcctagt caaaaaatgt  
 1081 gacacttggtg ccatagctga cttagggctg gctgtgaagc acgattctat catgaacact  
 1141 atagacatac cccagaatcc taaagtggga accaagaggt acatggctcc cgaaatgctt  
 1201 gatgatacaa tgaacctgag catctttgag tccttcaagc gagctgacat ctattcgggtg  
 1261 gggctgggtt actgggaaat agctcgaagg tgttcagttg gaggagtgtg tgaggagtac  
 1321 cagttgcctt actatgacat ggtgccttca gatccttcga tagaagaaat gaggaaggtt  
 1381 gtctgtgatc agaaactccg accaaacttc ccaaaccagt ggcaaagctg tgaagcgtg  
 1441 cgggtcatgg gaagaataat gcgtgagtg cgtgtacgca acggggcagc acgcctgaca  
 15 1501 gccctgcgag tgaagaagac catctcccag ctgtgtgtca aggaagactg taaagcttag  
 1561 ctctgcgtac agggcaaagg agagtgtca cagctttcct tcccgtctcc cgctttgtgt  
 1621 aaatgttttt gcttcttttc tgctgtgttt tattttagtt ctacctcaag gatgactcac  
 1681 tacagtgttg aagtgtccca aaggcagcat gtaaagataa cgctaaagt aagcatgggc  
 1741 aggtccttgac tttcccagtt tccatgttgt ccttagtttt atgtttaaag gggacacgac  
 20 1801 tcgttttttc tttttattta aggaggaaga tgttatgagt ataagataag ctacatacaa  
 1861 tggtaaaaaa aatattaaat tttgtgtgta cttgaacca gagcatccac aaatgagatg  
 1921 aagacatatt taaaatgtga actactgtat attatggtga aatggaactg taaaaattaa  
 1981 ctgggtcctt aaaaagttaa tttgtctcag aattttaaca tctaaagagt ggaaaggtgg  
 2041 agtaacaaag tgacaccacg ttctcaggac ttgctttctc tgcctatttg ccactctgca  
 25 2101 tcactacgac agctgtccct aagtgttcg aaaaccacct gagccattac tcctcctgat  
 2161 acataacaga ggtccatatg tctcaaacac tgtgaaatgg ccagagtgca cagagctctt  
 2221 gccagagctg gttacgaacc ctgtgacaac tagcactgtg aagtttgctg gccattatc  
 2281 tcccttgaat tcccatagag cgtttcttcc caagtgggaa gatcaatcat cagttataat  
 2341 atttttaaatt tagtcacata gtgaatgcat tgagtatttt gtaagaccat aatagatttt  
 30 2401 ctagaatttc agaaccaatc tattagaatt tcatcaaaga aatgtcatca ttatccatt  
 2461 cgataatgaa aaaaaatgtt ttctgaacat tgccaaatat tgaccaatta taaacatcca  
 2521 ttttctaaag atgtatgtca taagatgtca ttgcttttga ggttcatggg tggaaagaaa  
 2581 atgaaacacg gataaacacg gaaagaattt gaggagagt cagagaactc tcccgaagga  
 2641 gtgtaacccc ttactggtaa cagaagatca catcagtgat tacaaggcat tccctgctt  
 35 2701 ttaacaggga ccatgcatgg aagggtttgg aggatctacc tctattctcc atttttgaag  
 2761 ttcagaatat ttaacaggac accagatatt tctttcaaat gtcctatagg ctttagtgaa  
 2821 gatgggttaa ctcaacgtct ttcatgtgtc tgaacagtt tcccgaacct cattccttag  
 2881 aactttcatc tacctagaag ttgttagaat ggtgtgacac tctaacaagg tcttgaaaag  
 2941 tgatatgtta acatattcta ttctttgata gttctagcag caaagacaca ttccatggat  
 40 3001 atttttaagg gtaaatattt gattgtagcc tttctattat gagtaagggt aagttttact  
 3061 atctagcatc ttctaaagcc caacatggta aatgatgttg tgaagaatac gttatgtatc  
 3121 ccataagagg gtaacttgat gtctacatat cactatggag atatgctact cगतcatctt  
 3181 catttgacta agtggagata agagaaacaa aagctgacag ctgtggaggt tgcagtact  
 3241 tccattctca gtcattgctt catccattta taagcacagg gcaccccata acttacttgc  
 45 3301 cacaatatgg ggtacaagag ggaataacaa gatagatagg tcctgcccctc ggggatttta  
 3361 aagtctaatt tgggagtggg gaaagacatg tgagtatgag ggaagcgagt aaattgacag  
 3421 ataaaaatat cagtgggatg cctttgagtt ctatgtgaca atggcttcta gaatagggtc  
 3481 taaaatcact tccaatttgc aacacattta gaagggtatc ttatttagat attacagcct  
 3541 aattatatca tcagcatcaa aagggtgcgc tacgtgaatg gatgttaaga agaaccgtgc  
 50 3601 gaaggagag ctggcttatg tctattatct tatttcacag tttctacatg acggcaatgg  
 3661 aattatttac ttcaatgatt gtatccagat atccaatgaa aaaaacctac tctatcctgt  
 3721 tccattcacc actgttaggg aagtatatatt tctagatttt aaaagaggggt gttgtaacat  
 3781 ctttattttg gtcattatgt tttataattt atttcagggt atatacaaac ataacttata  
 3841 tttcttgggt ttagatttca cagtggagaa atctgcacat tggtagattt tgagactttt  
 55 3901 gttcattatt ctgttgctga actgtaatcc ctaagtgaat aatcttaaag agcagagcca  
 3961 cagaagtagt cagggaacaa tagaagttgt agtttttgct taactctcaa catagtctta  
 4021 catagatact gtcacttat aaaaaataga gtaattttta gagcataagg catattttact  
 4081 gtatatatct atgcagttca attttgtatt atgtgttttt gatggaatct tatgtaacac  
 4141 ttggctcctt taaattttgt tacgagcaaa aatatcttct taggttaggt gcaaaagtgc  
 60 4201 accattgctt tgctccttat ggatattttg gctctgggta taacatagag cttaggaatg  
 4261 ctattatgaa ttttttagtac tgtatgcaaa aaatgggtgg gtttatatac agcatcactt

4321 cttgaaatac agaacgcatt atttggtgtc tacatttgag catgaatttg ttgccttata  
 4381 agtcctgcgt tttgagtttg taatcagtag tgactctgtt gtatgcatcc ttccacgtct  
 4441 aaaatatttg gcatgtcaca tctagaattc ttaatttatg ttctgccatg agagttaagt  
 4501 gaaacatgac tgtcatgctc tattttaagc gcagcacttg cttttcatct ttatactttt  
 5 4561 caattaactt tgcattttta atttccataa ttgtatgaaa atagtaacct gatcgcaatg  
 4621 tctgaagagt ttcaaactcg cgtttatttt aaaatg

## SEQ ID NO: 16

Amino acid sequence of mouse ALK7 encoded by the DNA sequence shown in SEQ ID NO: 15.

10 MTPARGSALSALLLVALAADLAAGLKCVLLCDSSNFTCQTEGACWASVMLTNGKEQVI  
 KSCVSLPELNAQVFCSSNNVTKTECCFTDFCNNITLHLPTASPNAPRLGPTELTVVITV  
 PVCLLSIAAMLTIWACQDRQCTYRKTKRHNVEEALAEYSLVNAGKTLKDLIYDATASGSG  
 SGLPLLQRTIARTIVLQEIIVGKGRFGEVWHGRWCGEDVAVKIFSSRDERSWFREAEIYQ  
 TVMLRHENILGFIAADNKDNGTWTQLWLVSSEYHEQGSGLYDYLNRNIVTVAGMVKLALSIA  
 15 SGLAHLHMEIVGTQKPAIAHRDIKSKNILVKKCDTCAIADLGLAVKHDSIMNTIDIPQN  
 PKVGTKRYMAPEMLDDTMNLSIFESFKRADIYSVGLVYWEIARRCSVGGVVEEYQLPYD  
 MVPSDPSIEEMRKVVCQDKLRPNLPNQWQSCEALRMGRIMRECWYANGAARLTALRVKK  
 TISQLCVKEDCKA

## SEQ ID NO: 17

20 gi|20806128|ref|NM\_139090.1| Rattus norvegicus activin receptor-like kinase 7 (Alk7), mRNA

1 gggaggcccc gctgccacta gagccaaccg cgcacttcg aaggtgtcgc ggctggcctc  
 61 cctcccgccc gccccgggaa cttcaaagcg gcccgcgctg cgtgcccgtc tgggaccccg  
 121 aagccttgca ccgcccggg gtggccatga cccagcaag ccgctccgca ctgagcctgg  
 25 181 cctcctgct ggtggcactg gcctccgacc ttgcggcagg actgaagtgt gtgtgtcttt  
 241 tgtgtgattc ctcaaacttt acctgccaaa ccgaaggagc atgctgggcc tctgtcatgc  
 301 taaccaacgg gaaagaacag gtgatcaaat cgtgcgtgtc cctcccgaa ctaaagtctc  
 361 aggtctcttg tcacagttec aacaacgtga ccaagaccga atgttgcttc acagacttct  
 421 gcaacaacat cactctgcac cttcccacag catctccaga tgcccctaga cttggcccca  
 30 481 cagagctgac agttgttatt actgtacctg tttgcctcct gtccatcgca gccatgctaa  
 541 cgatatgggc ctgccaggac cgccagtgc catacaggaa gaccaagaga cacaatgtgg  
 601 aggaaccact ggcagagtac agccttgtca atgctggaaa aaccctcaa gatctgattt  
 661 atgatgccac tgcctcgggc tcaggatctg gcctgcctct tttggttcaa agaaccatcg  
 721 caaggacaat tgtacttcaa gaaatcgtag gaaaaggctg gtttggggaa gtgtggcacg  
 35 781 gaagatgggtg tggagaagat gtggctgtga aaatattctc ctccagagat gagagatctt  
 841 ggttcggtga ggcagaaatt tatcagacgg taatgctgag acatgagaat attctcggtt  
 901 tcatcgcggc cgacaacaaa gataatggaa cctggactca gctttggctt gtgtcagagt  
 961 atcacgagca gggctcctta tatgactatt tgaatagaaa catagtgacc gtggctggaa  
 1021 tggtaagtt ggcgctttca atagcgagt gtctggctca cctacacatg gagatcgtgg  
 40 1081 gcactcaagg taagcctgct attgctcacc gagatataaa gtcaaagaat atcttagtca  
 1141 aaaagtgtga cacttgctgc atagtctgact tagggctggc tgtgaaacat gattctatca  
 1201 tgaacactat agatataccc cagaatccta aagtgggaac caagaggat atggctcccg  
 1261 aaatgcttga tgatacaatg aacgtcaaca tctttgagtc cttcaagcga gctgacatct  
 1321 attcggtggg gctgggtttac tgggaaatag ctccgaagggtg ttcagttgga ggacttggtg  
 45 1381 aagagcacc gttgccttat tatgacatgg tgccttcaga tcttccata gaggaaatga  
 1441 ggaaggctgt ttgtgatcag aaactccgac caaatctccc aaaccagtgg caaagctgtg  
 1501 aggcgclccg ggtcatggga agaataatgc gtgagtgtg gtatgccaac ggggcagctc  
 1561 gcctgaccgc cctgcgcgtg aagaagacca tttctcagct gtgtgtcaag gaagactgta  
 1621 aggcctaagg cgcatacagg cgacgggaaa gccctcacca ctctctttca tgtctcctgc  
 50 1681 ttgtgtgtaaa tgtttctggt tcttttctgc tttgttttgt tttagttcta cctcaaagat  
 1741 gattcactac agtgttgaag tgtccaaagg cagcatgaaa agataactct aagcatgggc  
 1801 aggtcttgac tttcccagtt tccatgttgt ccgtactttt atttttaaaag gtgacactat

1861 tcatttttct ttttatttaa ggaggaagat gttat

SEQ ID NO: 18

Amino acid sequence of rat ALK7 encoded by the DNA sequence shown in SEQ ID NO: 17.

5 MTPASRSALSALLLVALASDLAAGLKCVCLLCDSSNFTCQTEGACWASVMLTNGKEQVI  
KSCVSLPELNAQVFCHSSNNVTKECCFTDFCNNITLHLPASPDAPRLGPTELTVVITV  
PVCLLSIAAMLTIWACQDRQCTYRKTKRHNVEEPLAEYSLVNAGKTLKDLYDATASGSG  
SGLPLLQRTIARTIVLQEI V GKGRFGEVWHGRWCGEDVAVKIFSSRDERSWFREAEIYQ  
TVMLRHENILGFIAADNKDNGTWTQLWL VSEYHEQGS LYDYLNRNIVTVAGMVKLALSIA  
10 SGLAHLHMEIVGTQGKPAIAHRDIKSKN I LVKKCDTCAIADLGLAVKHDSIMNTIDIPQN  
PKVGT KRYMAPEMLDDTMVN I FESFKRADIYSVGLVYWEIARRCSVGGGLVEEYQLPYD  
MVPSPDSIEEMRKVVCQKLRPNLPNQWQSCEALRMGRIMRECWYANGAARLTALRVKK  
TISQLCVKEDCKA

SEQ ID NO: 19

15 gi|21314629|ref|NM\_004054.2| Homo sapiens complement component 3a receptor 1  
(C3AR1), mRNA

1 actgtgggcta agtgtgggga ccagacagga ctctgtggaga catccagggtg ctgaagcctt  
61 cagctactgt ctctgttttt tgaagtttag caatggcgctc tttctctgct gagaccaatt  
121 caactgacct actctcacag ccatggaatg agccccagct aattctctcc atgggtcatte  
181 tcagccttac ttttttactg ggattgccag gcaatgggct ggtgctgtgg gtggctggcc  
20 241 tgaagatgca gcgacagtg aacacaattt ggttctctcca cctcaccttg gcggacctcc  
301 tctgtcgctt ctctctgccc ttctcgctgg ctcaactggc tctccaggga cagtggccct  
361 acggcagggt cctatgcaag ctcatccctt ccatcattgt cctcaacatg tttggccagt  
421 tcttctgctt tactgccatt agcctggatc gctgtcttgt ggtattcaag ccaatctggt  
481 gtcagaatca tcgcaatgta gggatggcct gctctatctg tggatgtatc tgggtggtgg  
25 541 cttttgtgat gtgcattcct gtgttcgtgt accgggaaat cttcactaca gacaaccata  
601 atagatgtgg ctacaaattt ggtctctcca gctcattaga ttatccagac ttttatggag  
661 atccactaga aaacaggctt ctgaaaaaca ttgttcagcc gcctggagaa atgaatgata  
721 ggtttagatc ttctcttctt caaacaatg atcatcttg gacagtcccc actgtcttcc  
781 aacctcaaac atttcaaaaga ccttctgcag attcactccc taggggttct cttgctttaa  
30 841 caagtcaaaa tctgtattct aatgtattta aacctgctga tgtgggtctca cctaaaatcc  
901 ccagtgggtt tcttattgaa gatcacgaaa ccagccact ggataactct gatgcttttc  
961 tctctactca tttaaagctg ttccctagcg cttctagcaa ttccttctac gagtctgagc  
1021 taccacaagg tttccaggat tattacaatt taggccaatt cacagatgac gatcaagtgc  
1081 caacaccctt cgtggcaata acgatcacta ggctagtggg gggtttctct ctgccctctg  
35 1141 ttatcatgat agcctgttac agcttcattg tcttccgaat gcaaaggggc cgcttcgcca  
1201 agtctcagag caaaaccttt cgagtggcgg tgggtggtgg ggctgtcttt cttgtctgtg  
1261 ggactccata ccacattttt ggagtcctgt cattgcttac tgaccagaa actcccttgg  
1321 ggaaaactct gatgtcctgg gatcatgtat gcattgctct agcatctgcc aatagttgct  
1381 ttaatccctt cctttatgcc ctcttgggga aagatttttag gaagaaagca aggcagtcca  
40 1441 ttcagggaat tctggaggca gccttcagtg aggagctcac acgttccacc cactgtccct  
1501 caaacaatgt catttcagaa agaaatagta caactgtgtg aaaatgtgga gcagccaaca  
1561 agcagggggt cttaggcaat cacatagtga aagtttataa gaggatgaag tgatatgggt  
1621 agcagcggaac ttcaaaaact gtcaaaagaat caatccagcg gttctcaaac ggtacacaga  
1681 ctattgacat cagcatcacc tagaaacttg ttagaaatgc aaattctcaa gccgcattcc  
45 1741 agacttgctg aatcggaaac tctgggggtt gggaccagc aagggcactt aacaaaccct  
1801 cgtttctgat taatgctaaa tgtaagaatc attgtaaaac ttagttctat ttctatccca  
1861 aactaagcta tgtgaaataa gagaagctac tttgttttta aatgatgttg aatatttgc  
1921 gatattttcca tcattaaatt tttccttagc attgtaaaaa aaaaaaaaaa aaaaaaaaaa  
1981 aaaaa

50 SEQ ID NO: 20



Amino acid sequence of human C3AR1 encoded by the DNA sequence shown in SEQ ID NO: 19.

5 MASPSAETNSTDLLSQPWNEPPVILSMVILSLTFLGLPGNGLVLWVAGLKMORTVNTIW  
 FLHLTLADLLCCLSLPFLAHLALQGQWPYGRFLCKLIPSIIVLNMFASVFLTAISLDR  
 CLVVFKEPIWCQNHNRVGMACSIICGCIWVAVFVMCIPVFVYREIFTTDNHNRCGYKFGLSS  
 SLDYPDFYGDPLENRSLENIVQPPGEMNDRDPSSFQTNHDPWTVPTVFQPTFQRPASD  
 SLPRGSARLTSQNLYSNVFKPADVVSFKIPSGFPIEDHETSPLDNSDAFLSTHLKLFPSA  
 SSNSPYESELPGGFQDYNLGQFTDDQVPTPLVAITITRLVVGFLPSVIMIACYSFIV  
 10 FRMQGRGFAKSQSKTFRVAVVVAVFLVCWTPYHIFGVLSLLTDPETPLGKTLMSWDHVC  
 IALASANSFCNPFYLLGKDFRKKARQSIQGILEAAFSEELTRSTHCPSNNVISERNST  
 TV

SEQ ID NO: 21

gi|6753223|ref|NM\_009779.1| Mus musculus complement component 3a receptor 1 (C3ar1), mRNA

15 1 gaattccatc tcagtgtgct tgactgagcc atggagtctt tcgatgctga caccaattca  
 61 actgacctac actcacggcc tctgtttcaa cccaagaca ttgctccat ggtcattctt  
 121 ggtctcactt gtctattggg actgctagcc aatgggctgg tgctgtgggt agctggcgta  
 181 aagatgaaga cgaccgtgaa cacagtctgg ttccctccatc tcaccctggc cgatttcctc  
 241 tgctgcctct ccttgccctt ctcttggtt caccgtatc tccaaggaca ctggccctat  
 20 301 ggcttggtcc tgtgcaaac tatcccatcc atcattattc tcaacatgtt tgccagtgtc  
 361 ttctgtctta ctgccattag cctggaccga tgtctgatag tacataagcc aatctgggtc  
 421 cagaatcatc gaaacgtgag aaccgccttc gccatctgtg gatgtgtctg ggtggtagcc  
 481 tttgtgatgt gtgtgcccgt atttgtatac cgtgatctgt tcattatgga caatcgcagt  
 541 atatgtagat ataattttga ttccctccagg tcatatgatt attgggacta cgtgtacaaa  
 25 601 ctaagtctac cagaaagcaa ttctactgat aactccactg ctacagtaac tggacatatg  
 661 aatgacaggt cagctccttc ctctgtacag gcaagggtt acttttggtg agttaccact  
 721 gccctccagt cacagccatt cctaagact cctgaagact cattctctct agattcagca  
 781 aaccaacaac ccattatgg tggaagcct cctaagtcc tcacagccgc cgtaccagc  
 841 gggtttccctg ttgaagatcg taaatccaat acactgaacg ctgacgcttt tctctctgct  
 30 901 cacacagaac ttttccctac tgcttctagt ggtcatttat accctatga tttccagggg  
 961 gattatgttg accaattcac gtatgacaat catgtgccga caccgctgat ggcaataacc  
 1021 atcacaaggc tgggtgggtgg cttcctgggtg ccgtttttca tcatggtaat ttgttacagc  
 1081 ctcatcgtct tcagaatgag aaaaaccaac ttcaccaagt ctcggaacaa aacctttcgg  
 1141 gtggctgtgg ctgtgggtcac tgtctttttt atctgtctgga ctccatacca tctgtctgga  
 35 1201 gtctctctat tgattactga tccagaaagt tccctggggg aagctgtgat gctctgggac  
 1261 cacatgtcca ttgcttttagc atctgccaat agttgcttca accctttcct gtatgccctc  
 1321 ttggggaaag acttttaggaa gaaagcaaga cagtctataa agggcattct ggaagcagcc  
 1381 ttcagcgaag agctcacgca ctctaccaac tgtaccaag acaaagcctc ttcaaaaaga  
 1441 aacaatatga gtacagatgt gtgaagatgt ggccctggga acctaagcag agttctcagg  
 40 1501 tgaacagtga tggatgacat gtgagcagga cacttttagc aatttggcga ctctcagaga  
 1561 aaggtctctt attgacatca gcatcatttg aaacattaa agatgcaaaa tttcaagccc  
 1621 catccagat gtgttgactc agaattctctg gcccatggga ccagtgtttt aacaggcctt  
 1681 cttgtttcca tcagtgttaa gttttacctc atttggttta gtctattccc atccctgact  
 1741 acaccatgtg caatgaataa ctttttcac tgttttcagt attctttttt tttccttagc  
 45 1801 atcatctaaa cttctagttt gcatggagg ctgctcttat tgttctgaat ggaagatatt  
 1861 catttattgt acagttttgt ggtggtgaca agtgattttt aagtggggaa agagacacag  
 1921 taagaaaaga tctatgaaag cagggagtgt tgagttagag ttgacagaa cacagtgcc  
 1981 aatgccaccc actaaaagca attccagtg tcatgtgagc aagtgtgac tggagctct  
 2041 agatacacat aaacactttc ctactctctg agtggttttag aagttgtagc ttggagctct  
 50 2101 ctcttccctg tgtattctgc ttgttcatag agccatctat gttccatttc ttttgtttg  
 2161 ttggtttggc ttctgtgtt ggttgggtt ttggttgggt ggttgggtt ttggttgggt  
 2221 gcttgggtt ttggttgggt ggttgggtt ttgcttgggt gcttgggtt ctgttttttc  
 2281 aagacagtag ccctggctgt cctgaaactc attctgtaga ccaaactggc ctgaaactca  
 2341 gagatcagtg tgctctgccc tccaaggagt cttcttaagg tcttgaattc ctttctgttt  
 55 2401 tccacacagg gccttgccat ctctgctgg gattaggcgt gtgcaggtgt gcagtacgta

```

2461 cctggcccca tcgtcttcgg ctttttacat tagactttgt atctcacata gaataaaaca
2521 atatccta at gggatgatgg gtggaacttc aggatcaatg catgcttcaa cagcaatatc
2581 aaaagatttt aaacatgagt ctaggatctt tacatgacaa aagtgtcagg cacactctct
2641 tttttttctt tctttttctt ttcttttttt tttttttcaa ttaatgtttg cctggaccat
5 2701 gctaataata agtactgtat gcagccaatt ttaaaggaac taattaaaaa tgaaataaaa
2761 aagggtgggt tctccagtgc ccagtagttc ccatctctcc ctactttgca agaattgtag
2821 ctgttccttc tgatatattac catgggtgtgc ctaaaacacc tcttatgtcg ctgatctttg
2881 tagacaggaa gaaaaggctg tcttgatttt gatttacttt acctacattt tcatccagca
2941 cccagtggtt gactgggtatc acacattaat ctgttcgtga tcatttttct cttccctttg
10 3001 ttgtcgggtt ttgcttgcc tctgtgccatc atctcagtgt agcacacaca atctagcttc
3061 agactcccag gtctcttggt atagtaacca gatgagcacc accaccatcc ccacccccag
3121 ctgaccacac aattttctcg tcaatacatt tcaatagcat acagcttttc tctgtcctct
3181 catttttcat gctgacatat attaattgta catattaata aggttccacg tgggtgatcc
3241 atccatgcac acaacattca cactttcacc cccagcctct tctttatcat tattctactt
15 3301 actgattgac tttaaaaaaa aaaaaaaaaa

```

SEQ ID NO: 22

Amino acid sequence of mouse C3AR1 encoded by the DNA sequence shown in SEQ ID NO: 21.

```

20 MESFDADTNSDLHSRPLFQPDIASMVILGLTCLLGLLGNGLVLWVAGVKMKTTVNTVW
FLHLTLADFLCCLSLPFLSLAHLILQGHWPYGLFLCKLIPSIILNMFASVFLLTALSLDR
CLIVHKPIWCQNHNRNVRTAFATCGCVWVAFVMCPVFVYRDLFIMDNRSICRYNFDSSR
SYDYWDYVYKLSLPESNSTDNSTAQLTGHMNDRSAPSSVQARDYFWTVTTALQSQPFLTS
25 PEDSFSLDSANQQPHYGGKPPNVLTAAVPSGFPVEDRKSNTLNADAFLSAHTELFPTASS
GHLYPYDFQGDYVDQFTYDNHVPPLMAITITRLVVGFLVPPFIMVICYSLIVFRMRKTN
FTKSRNKTFRVAVAVTVFFICWTPYHLVGVLLITDPESSLGEAVMSWDHMSIALASAN
SCFNPFYALLGKDFRKKARQSIKILEAAFSEELTHSTNCTQDKASSKRNMMSTDV

```

SEQ ID NO: 23

gi|14091739|ref|NM\_032060.1| Rattus norvegicus complement component 3a receptor 1 (C3ar1), mRNA

```

30 1 1 tgtcagagtaa gcactcagca actagtccaa catcctccac ggcggtgga ccagatagga
61 tttgttggtg gctcacagat catcagtcct ggagccttct gatccacact cagtgtctct
121 gactgagcaa tggagtcttt cactgctgac accaactcaa ctgacctaca ctcacggcct
181 ctgttttaaac cccaagacat tgctccatg gtcattctca gtctcacgtg tttattggga
241 ctgccaggca atgggctggt gctgtgggtc gctggcgtaa agatgaagag gaccgtgaac
35 301 acagtctggt tcctccacct caccctggcc gatttcctct gctgcctctc cttgcccttc
361 tccgtggctc acctgattct ccgaggacac tggccctatg gcttgttctt gtgcaaaactt
421 atcccgtccg tcattatcct caacatgttt gccagtgtct tcctgcttagc tgccattagc
481 ctggaccgct gtctgatggt gcacaagccg atctgggtgcc agaaccatcg gagcgtgaga
541 acagccttcg ccgtctgttg atgcgtctgg gtggtaacct ttgtgatgtg tataccggta
40 601 tttgtatacc gtgatctgct cgttgtggat gattacagtg tatgtggcta taactttgat
661 tcctccaggg catatgatta ttgggactac atgtacaatt cacatctacc agaaatcaac
721 cctcctgaca actccactgg acatgtggat gacaggacag ctcttctctc ctctgtacca
781 gcaagggacc tttggacagc taccactgct ctccagtcac aaacgttcca cacatctcct
841 gaagaccat tctctcaaga ctcagcaagt cagcaacccc attatggggg gaagcctcct
45 901 actgtgtctc tagccaccat acccggcggt tttcctgttg aagatcataa atccaacaca
961 ctgaacactg gtgtttttct ctctgcccac acagagccgt ccctcactgc ttctagcagt
1021 cctttatacg cccatgattt cccggatgat tattttgacc aactgatgta tggcaatcac
1081 gcgtggacac ctcaggtggc aatcaccatc tcgaggctgg tgggtgggctt cctggtgccc
1141 tttttcatca tgataacctg ttacagcctc atcgtcttca gaatgcgaaa aaccaacctc
50 1201 accaagtcca ggaacaaaac ccttcgggtg gccgtggcgg tggtcaccgt cttttttgtc
1261 tgctggattc cataccatat tgttggaaatc ctgttagtga ttactgacca agaaagcgcc
1321 ttgagagaag ttgtgctacc ctgggaccac atgtccattg ctttagcatc tgccaatagt

```

1381 tgtttcaacc ctttcttcta tgccctcttg ggaaaagact tcaggaagaa agcaagacag  
 1441 tccgtaaagg gcattctgga agcagccttc agcgaagagc tcacgcactc taccagttgt  
 1501 acccaggaca aagccccctc aaaaagaaac catatgagta cggatgtgtg aagatgctgc  
 1561 actggggacc taagcagatg ttctgaggtg aatactggtg atgggtgacc tgtgagcggg  
 5 1621 acaccttaga cagcctggcc accctcagag aaaggtctct tattgacatc agcatcattt  
 1681 gaaaacacta aaggcacaaa atttcaagcc ccatcccaga tgtttgactc cgaatctctg  
 1741 gcccgtgggg accaatgtct taacaggccc ccttgtttcc accaatgtta agttttattc  
 1801 aactcatttt attcccatcc ctgaatcggc catgtgcaat gaataacgtc ttcatctgtt  
 1861 tccagtatta atctcttctt gcatagcatc atttaaatcc tttagtttgt gcgggaggct  
 10 1921 gctattgtcc agagtgaag atattttttt attgaacatt tgtgtggtg tagcagtgat  
 1981 tttttagtgg ggaggcaggg gagaaagaca cagaaataaa aggtttttgg aaaaaaaaaa  
 2041 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa gaaaaaaaaa aaaaaaaaaa aaa

SEQ ID NO: 24

15 Amino acid sequence of rat C3AR1 encoded by the DNA sequence shown in SEQ ID NO: 23.

MESFTADTNSDLHSRPLFKPQDIASMVILSLTCLLGLPGNGLVLVWAGVKMKRTVNTVW  
 FLHLTLADFLCCLSLPFSVAHLILRGHWPYGLFLCKLIPSVIILNMFASVFLTLAISLDR  
 CLMVHKPIWCQNHSRVRTAFVCGCVWVVTVMCIPVFVYRDLLVDDYSVCGYNFDDSR  
 AYDYWDYMYNSHLPEINPPDNTSGHVDDRTAPSSSVPARDLWTATTALQSQTFTSPEDP  
 20 FSQDSASQPHYGGKPPVTIATIPGGFPVEDHKSNTLNTGAFLSAHTEPSLTASSSPLY  
 AHDFPDDYFDQLMYGNHAWTPQVAITISRLVVGFLVPPFFIMITCYSLIVFRMRKTNLTKS  
 RNKTLRVAVAVVTFFVCWIPYHIVGILLVITDQESALREVLPWDHMSIALASANSFCN  
 PFLYALLGKDFRKARQSVKGILEAAFSEELTHSTCTQDKAPSKRNMSTDV

SEQ ID NO: 25

25 gi|34878691|ref|NM\_005795.2| Homo sapiens calcitonin receptor-like (CALCRL), mRNA

1 gaacaacctc tctctctcca gcagagagtg tcacctcctg ctttaggacc atcaagctct  
 61 gctaactgaa tctcatccta attgcaggat cacattgcaa agctttcact ctttcccacc  
 121 ttgcttggtg gtaaatctct tctgcggaat ctcagaaagt aaagttccat cctgagaata  
 181 tttcacaaag aatttcctta agagctggac tgggtcttga cccctgaatt taagaaattc  
 30 241 ttaaagacaa tgtcaaatat gatccaagag aaaaatgtgat ttgagtctgg agacaattgt  
 301 gcatatcgte taataataaa aaccataact agcctataga aaacaatatt tgaaagattg  
 361 ctaccactaa aaagaaaact actacaactt gacaagactg ctgcaaaact caatttgtca  
 421 accacaactt gacaagggtg ctataaaaaca agattgctac aacttctagt ttatgttata  
 481 cagcataattt cattttggct taatgatgga gaaaaagtgt accctgtatt ttctggttct  
 35 541 cttgcctttt tttatgattc ttgttacagc agaattagaa gagagtcctg aggactcaat  
 601 tcagttggga gttactagaa ataaaatcat gacagctcaa tatgaatgtt accaaaagat  
 661 tatgcaagac cccattcaac aagcagaagg cgtttactgc aacagaacct gggatggatg  
 721 gctctgctgg aacgatgttg cagcaggaac tgaatcaatg cagctctgcc ctgattactt  
 781 tcaggacttt gatccatcag aaaaagttac aaagatctgt gaccaagatg gaaactgggt  
 40 841 tagacatcca gcaagcaaca gaacatggac aaattatacc cagtgtaatg ttaacacca  
 901 cgagaaagtg aagactgcac taaatttggt ttacctgacc ataattggac acggattgtc  
 961 tattgcatca ctgcttatct cgcttggcat attcttttat ttcaagagcc taagtggcca  
 1021 aaggattacc ttacacaaaa atctgttctt ctcatattgt tgtaactctg ttgtaacaat  
 1081 cattcacctc actgcagtgg ccaacaacca ggccttagta gccacaaatc ctgttagttg  
 45 1141 caaagtgtcc cagtccatc atctttacct gatgggctgt aattactttt ggatgctctg  
 1201 tgaaggcatt tacctacaca cactcattgt ggtggccgtg tttgcagaga agcaacattt  
 1261 aatgtgggat tattttcttg gctggggatt tccactgat tctgcttgta tacatgccat  
 1321 tgctagaagg ttatattaca atgacaattg ctggatcagt tctgataccc atctcctcta  
 1381 cattatccat ggcccaattt gtgctgcttt actgggtgaat ctttttttct tgttaaatat  
 50 1441 tgtacgcgtt ctcatcacca agttaaaagt tacacaccaa gcggaatcca atctgtacat  
 1501 gaaagctgtg agagctactc ttatcttggt gccattgctt ggcattgaat ttgtgctgat  
 1561 tccatggcga cctgaaggaa agattgcaga ggaggtatat gactacatca tgcacatcct  
 1621 tatgcacttc cagggtcttt tggctctctac cattttctgc ttctttaatg gagaggttca

```

1681 agcaattctg agaagaaact ggaatcaata caaaatccaa tttggaaaca gcttttccaa
1741 ctcagaagct ctctgtagtg cgtcttacac agtgtcaaca atcagtgatg gtcagggtta
1801 tagtcatgac tgccttagtg aacacttaaa tggaaaaagc atccatgata ttgaaaatgt
1861 tctcttaaaa ccagaaaatt tatataattg aaaatagaag gatgggtgtc tcactgtttt
5 1921 gtgcttctcc taactcaagg acttggaacc atgactctgt agccagaaga cttcaatatt
1981 aaatgacttt ttgaatgtca taaagaagag ccttcacatg aaattagtag tgtgttgata
2041 agagtgtaac atccagctct atgtgggaaa aaagaaatcc tggtttgtaa tgtttgtcag
2101 taaatactcc cactatgcct gatgtgacgc tactaacctg acatcaccaa gtgtggaatt
2161 ggagaaaaagc acaatcaact tttctgagct ggtgtaagcc agttccagca caccattgca
10 2221 tgaattcaca aacaaatggc tgtaaaacta aacatacatg ttgggcatga ttctaccctt
2281 attgccccaa gagacctagc taaggtctag aaacatgaag ggaaaattag cttttagttt
2341 taaaactctt tatcccatct tgattggggc agttgacttt tttttgccc agagtgcctg
2401 agtctctttt gtaactaccc tctcaaatgg acaataccag aagtgaatta tccctgctgg
2461 ctttcttttc tctatgaaaa gcaactgagt acaattgtta tgatctactc atttgcctgac
15 2521 acatcagtta tatcttctgg catatccatt gtggaaactg gatgaacagg atgtataata
2581 tgcaatccta cttctatatac attaggaaaa catcttagtt gatgtacaa aacacctgtt
2641 caacctcttc ctgtcttacc aaacagtggg aggggaattcc tagctgtaaa tataaatttt
2701 gtcctctcca tttctactgt ataaacaaat tagcaatcat tttatataaa gaaaatcaat
2761 gaaggatttc ttattttctt ggaattttgt aaaaagaaat tgtgaaaaat gagcttgtaa
20 2821 atactccatt attttatttt atagtctcaa atcaaataca tacaacctat gtaattttta
2881 aagcaaatat ataatgcaac aatgtgtgta tgtaaatatc tgatactgta tctgggctga
2941 ttttttaaat aaaatagagt ctggaatgct aaaaaaaaaa aaaa

```

SEQ ID NO: 26

25 Amino acid sequence of human CALCRL encoded by the DNA sequence shown in SEQ ID NO: 25.

```

MEKKCTLYFLVLLPFFMILVTAELEESPEDSIQLGVTRNKIMTAQYECYQKIMQDPIQQA
EGVYCNRTWDGWLWCWNDVAAGTESMQLCPDYFQDFDPSEKVTKICDQDGNWFRHPASNRT
WTNYTQCNVNTHEKVKTALNLFYLTIIHGHLISIASLLISLGIFFYFKSLSCQRITLHKNL
FFSFVCNSVVTIIHLTAVANNQALVATNPVSCVKVSQFIHLYLMGCNYFWMLCEGIYLHTL
30 IIVAVFAEKQHLMWYYFLGWGPFPLIPACIIHAIARSLLYNDNCWISSDTHLLYIIHGPICA
ALLVNLFFLLNIVRVLITKLKVTHQAESNLYMKAVRATLILVPLLGIIEFVLIPWRPEGKI
AEEVYDYIMHILMHFQGLLVSTIFCFNGEVQAILRRNWNQYKIQFGNSFSNSEALRSAS
YTVSTISDGPYSHDCPSEHLNGKSIHDIENVLLKPENLYN

```

SEQ ID NO: 27

35 gi|9055257|ref|NM\_018782.1| Mus musculus calcitonin receptor-like (Calcrl), mRNA

```

1  tgagtctgga gacaattgtg tatgtatact tttcttaaga tattaataaa caaatccaag
61  gtcacagggt gcttattgat agaagagaaa caatacggat agaagagaaa tcagaaaatt
121 gcttatgatt gacaagaaca gctgcagcag ctacctagct tgaacataca gcacatttca
181 tttggactct aataatggat aaaaagcata tactatgttt tctggttctc ttgcctctta
40 241 atatggctct catctcagca gagtccgaag aaggcgtgaa ccaaacagac ttgggagtca
301 ctagaacaca gatcatgacg gctcaatatg aatgttacca gaagatcatg caggacccca
361 ttcaacaagc agaaggcctt tactgcaata ggacctggga cggatggcta tgctggaatg
421 acgttgacgc agggacggaa tcaatgcagt actgccctga ctattttcag gattttgatc
481 cttcagagaa ggttacaaag atctgtgacc aagatggaca ctggtttcgg catccggata
45 541 gtaatagaac atggaccaac tacacctgtg gtaataacag cacgcatgag aaagtgaaga
601 cagccctgaa tctgttctac ctaactataa ttggacatgg attatctatt gcactcttga
661 tcatctctct catcatattt ttttacttca agagcctaag ttgccaacgg atcacattgc
721 ataaaaacct gttcttttca tttatttgta attcgattgt aacaatcatc cacctcacgg
781 cagtggccaa taaccaggcc ttagtggcca caaatcctgt gagctgcaaa ggcgtcagc
50 841 ttatccatct ctacctgatg ggctgtaact acttctggat gctctgtgaa ggcgtttacc
901 tgcacacact catcgtgggt gctgtgtttg cggagaagca gcacttgatg tggattattt
961 ttctcggctg ggggtttcct ctgcttctct cctgcatcca cgccattgcc agaagcttgt
1021 attacaacga caattgctgg atcagctcag acactcatct cctctacatt atccatggtc

```

```

1081 cgatttgtgc tgccttgttg gtaaatctct ttttccctatt aaatattgta cgtgttctca
1141 tcaccaagtt gaaagttaca caccaagtgg aatccaatct ctacatgaaa gccgtaagag
1201 ctactctcat cttggtagca ctacttggca ttgaatttgt gctttttccg tggcgccctg
1261 aaggaaaggt tgcagaggag gtgtatgact atgtcatgca cattttgatg cactttcagg
5 1321 gtcttttggg ggctactatt ttctgcttct ttaatggaga gggtcaagca attctgagaa
1381 gaaattggaa ccagtataaa atccaatttg gaaatggctt tcccactct gatgctctcc
1441 gcagtgcac ctcacacagt tcaacaatca gtgacatgca aggttacagc catgactgcc
1501 ccaactgaaca cttaaatgga aaaagcatcc aggatattga aaatgttgcc ttaaaatcag
1561 aaaataccta tgatctagt atgtgaaaat atataaaatc acacacttga acctatgatt
10 1621 ttatagccaa aagattggat agccagcaaa tggatttctg gaataccatg aagaaagccc
1681 tcaaatgaaa tgggaattgt atggataaat gtttaacaac cctctctcta tgggggagaa
1741 aagcctcaat ttatattgtt ggccagtaac tactcctacc ataactgatt tcaagttacc
1801 aacctgacat cactgaatgt gtaattggaa agaaaataag cacaacaac tcccaggagc
1861 tgacatgttc tggaaacctgc acagcactgc atgccctcag agatggaagt ctgtgacac
15 1921 tctcaccac actgaccaga actctgctct tctcatctga ggagacctca ctcgactta
1981 cagacatgaa aggaaagtct gggttttgggt ataaaactcc tcatcctttt gtcatttggg
2041 gatagtgaat gggttgtcca gaaacacttt aacctctttt ataacttctc ttgtagacta
2101 gacaaagagt tcattatctc tgtggtctca tgttactttc tttaagaaaa gcaactgagc
2161 agaattccca tgggcttctt agttgctgtt acatattgtt gtatcctgtg gtatattgcac
20 2221 tctgaattcac tagggagaca tattggtaga agctacacct tgtcggtctc ctctctctct
2281 ctctctctct ctctctctct ctctctctct ctctctcggc aagaatagga ggaaatcagt
2341 tttctggaat ctgtaaatag aaactttggc ctttccattt ctactgtgta gataagtaag
2401 tgatcaacca ttttacatga agggaaatcaa tgaaggattt ctttttatct tgggaatctg
2461 agagagagag agagagagag agagagagag agagagagag agagagagag agagagagag
25 2521 aggttgtaca aatgaaactg taagtactct gtaattttat tttattgttt caattcaaaa
2581 cacaacctag gtaactttta aagcagatag atgatgcaac attatatgca acttagtacc
2641 tgatactgtg tctgggctga tttattttat agtaaaatag aatcatgaat tctatatgtg
2701 gtaaatatth taaggacaaa cagataccag catcagaagt ttgagaacta aaaacaaacc
2761 ccagaaatat caatgataag atatgaaata tttattttaa aatgtaaagc cataatttta
30 2821 caagcatgat atatttgggt tgctgacata ctgcttccaa tatgacatat tccaatttga
2881 tgacaactca ttcaaattca gaaattatga acaactactg taagaagtgc tagtctgcta
2941 catttggggg ctttaattca ctaaattata gctttactga tttttactaa atgaagatac
3001 atgcttttga aaaagaccca gggtattatt tttttatgaa ccacagagga attttctcta
3061 tgaataaaca ctgtttcagt atttcagtag actactgtgg atgataatc aagccacggg
35 3121 aagacttgaa cttaccactc attgaggatg aagtgtcaga tatttaagta ataattaaat
3181 taagacagaa gaagattttg cttgtgagtt taagtacat tttaaatgga ttagtttttt
3241 ccaatatgtc ttgggttggg gtttctgaag aacaagattg acttcagtta gtaaaagtca
3301 ctctggggac agttacatga gtccatgttt tttaaagtgt gtctgtcagt ttacagtgga
3361 tgagaggagt gactcagccc tcactcattg tcttaagcta tagaagagaa taaattaaaa
40 3421 agttgtttga aactgtcaag gaaaagaatt caaattatca agcactgatg tattacatga
3481 tctgacagac atatttgtaa tttctttttt tttttttctt tctttttttt tatatttga
3541 attttctgaa ctttcagctg cttaaaccgt acaaaccttc cttttaatac cttcagactt
3601 taggtattga ttattcatat cttcgctgac agtgttcgtg taattctgca ttcttctaaa
3661 caaactgtca agtattaaca ataaatggta gtacagaatg ccgaattacc agagaacata
45 3721 cttaaaaggt attgtaattt taagttttat tattttttta ctataccgag tgtagctgag
3781 atttttttac tgccaatatc tactgtggtg tcattctcag ccatagcagt ctaagattag
3841 tcagtattag ttttgaata gtatttttga agctggattt ctctaaagct gaagtccaa
3901 ttaaaatatt tctataatgc ttgtagaaaa aaaatccaaa aattttgaaa gaaaacttct
3961 ctggtgccta aaaattaact tggggagttt ttagctttta aaagttgcat ctattatttt
50 4021 tctctaatac tttgttagag tcatggctga cactttttaga agcgactttg tatttttttg
4081 ttgataaacc agccttctct gaggttagtg ttttttcaa tataacaaac tgtgtagggg
4141 ttaggagtag tcagtgttat acttactgtt tagatatttt aatgtgaatc tattttaatt
4201 tcctttgtgt ttttaattgc tcatagaaac attatatgta aaacatactc aagttcactt
4261 aaaagaactc atattttatc gacataaaat ctctacttag gttttatgag aatattaaag
55 4321 ttgtaagtta tcaactactt tcacgcaccc tccctggact atagagcaat tgaattcctc
4381 acaaaattct tgcagcaaat aaagaaatct tttgcaaaat aaaaaaaaaa aaaa'aaaaa
4441 aaaaaaa

```

SEQ ID NO: 28

Amino acid sequence of mouse CALCRL encoded by the DNA sequence shown in SEQ ID NO: 27.

MDKKHILCFLVLLPLNMALISAEESEGVNQTDLGVTRNKIMTAQYECYQKIMQDPIQQAE  
 GLYCNRTWDGWLWCNDVAAGTESMQYCPDYFQDFDPSEKVTIKCDQDGHWFHRPDSNRTW  
 5 TNYTLCNNSTHEKVKLTALNLFYLTIIHGGLSIALSIISLIIFFYFKSLSCQRITLHKNLF  
 PSFICNSIVTIIHLTAVANNQALVATNPVSKVSQFIHLYLMGCNYFWMLCEGVYLHTLI  
 VVAVFAEKQHLMWYYFLGWGFLLPACIHAIARSLYNDNCWISSDTHLLYIIHGPICAA  
 LLVNLFFLLNIVRVLITKLKVTHQVESNLYMKAVRATLILVPLLGI FVLFPWRPEGKVA  
 10 EEVDYVMHILMHFQGLLVATIFCFPFNGEVQAILRRNWNQYKIQFGNGFSHSDALRSASY  
 TVSTISDMQGYSHDCPTEHLNGKSIQDIENVALKSENTYDLVM

SEQ ID NO: 29

gi|6978586|ref|NM\_012717.1| Rattus norvegicus calcitonin receptor-like (Calcrl), mRNA

1 gggaaactctt ttctactatc tcagaaaatc aaattccctc ctgagactat tcacacagaa  
 61 tttcccttagg agccgtgagg ggcctatga tgatacacca acatcttggtg ctaccaactg  
 15 121 tcaaaaacttt ggatgggtcta acactcagggt catgaccctt gtaattaaaga aattctcgaa  
 181 gacaatattg agcatgattc aggagaaaaat gtgatttgag tctggagaca attgtgcatg  
 241 tatatctttt ctttaagatct taaaaagaaa gagaaaacaa atccagggtt gtcggtcact  
 301 gattgataga agagaaaaca atattcagac gattgcttat gattgacaag agcagctgag  
 361 gccgctaccc agcttgaaca tatcgacacac ttcagttgga ctctaatgat ggataaaaag  
 421 tgtacactat gttttctggt tctcttgctt cttaatatgg ctctcatcgc agcagagtgc  
 481 gaagaaggcg cgaaccaaac agacttgagg gtcactagga acaagatcat gacggctcag  
 541 tatgaatggt accaaaagat catgcaggat cccattcaac aaggagaagg cgttactgac  
 601 aacagaaact gggacggatg gctatgctgg aatgacgttg cagcaggaac cgagtcaatg  
 661 cagtactgcc ctgattactt tcaagatttt gatccttcag agaagggtac aaagatctgt  
 721 gaccaagatg gaaactgggt cagacatcca gatagtaaca ggacatggac aaactacacc  
 781 ttgtgtaaca acagcacgca tgagaaagtg aagacagcac tgaatttggt ctacctact  
 841 ataattggac atggattatc tattgcctct ctgatcatct cactcatcat atttttttat  
 901 ttcaagagcc taagttgcca acggattaca ttgcataaaa acctgttctt ttcatttggt  
 961 tgaattcga ttgtgacaat cattcacctc acggcagtggt ccaataacca ggccttagtg  
 1021 gccacaaatc ctgtgagctg caagggtgtc cagttcattc atctttacct gatgggtgtg  
 1081 aactactttt ggatgctctg tgaaggcatt tacctgcaca cactcattgt ggtggctgtg  
 1141 tttgcagaga agcagcactt gatgtggtat tattttcttg gctgggggtt tctctgctt  
 1201 cctgcctgca tccatgccat cgcagaaagc ttgtattaca atgacaactg ctggatcagc  
 1261 tcagacactc atctcctcta catcatccat ggteccattt gtgctgcttt actggtaaat  
 1321 ctctttttcc tattaatat tgtacgtgtt ctcatcaca agttgaaagt tacacacca  
 1381 gcagaatcca atctctacat gaaagctgta agaaccactc tcatcttggt accactactt  
 1441 ggcatgtaat ttgtgctttt tccatggcgg cctgaaggaa aggttgctga ggaggtgtat  
 1501 gactatgtca tgcacattct catgcactat cagggtcttt tgggtgtctac aattttctgc  
 1561 ttctttaacg gagagggttca agcaattctg agaagaaatt ggaaccagta taaaatccaa  
 1621 tttggcaatg gcttttccca ctctgatgct ctccgcagcg catcctatac ggtgtcaaca  
 1681 atcagcgatg tgcaggggta cagccacgac tgccccactg aacacttaaa tggaaaaagc  
 1741 atccaggata ttgaaaatgt tgccttaaaa ccagaaaaaa tgtatgatct agtcatgtga  
 1801 aaatacagaa atctcacttg ttttgtgtca cttgaacctt tgattttata gtcggaagat  
 1861 ttaatagcaa atggggtttct tgaataccat gaagaaagcc ctcaaatgaa atgagaattg  
 1921 tgtggataaa tgtttaacga ccctctctct atgggggaga aaagcctcaa tttttattgt  
 1981 tggccagtaa atactcctac catagctgat ttcaagttac caacctgaca tcaccgaatg  
 2041 tgggaattgga aagaaaataa gcacaatcaa ctcttgggag ctgacatgag tcgctc

SEQ ID NO: 30

50 Amino acid sequence of rat CALCRL encoded by the DNA sequence shown in SEQ ID NO: 29.

MMDKKCTLCFLFLLLLNMLIAAESEEGANQTDLGVTRNKIMTAQYECYQKIMQDP IQQG  
 EGLYCNRTWDGWL CWNDAAGTESM QYCPDYFQDFDPSEKVT KICDQDGNWFRHPDSNRT  
 WNTYTL CNNSTHEKVK TALNLFYLT IIGHGLS IASLIISLI IFFYFKSLSCQRITLHKNL  
 FFSFVCNSIVTIIHLTAVANNQALVATNPVSKVSQPIHLYLMGCNYFWMLCEGIY LHTL  
 5 IVVAVFAEKQHLMWYFLGWGFLLPACIHAIARSLYNDNCWISSDTHLLYI IHGPICA  
 ALLVNLFFLLNIVRVLITKLKVTHQAESNLYMKAVRATLILVPLLGIEFVLF PWRPEGKV  
 AEEVYDYVMHILMHYQGLLVSTIFCFFNGEVQAILRRNWNQYKIQFGNGFSHSDALRSAS  
 YTVSTISDVQGYSHDCPTEHLNGKSIQDIENVALKPEKMYDLVM

SEQ ID NO: 31

10 gi|22538799|ref|NM\_005408.2| Homo sapiens chemokine (C-C motif) ligand 13 (CCL13), mRNA

1 aaaaggccgg cggaacagcc agaggagcag agaggcaaag aaacattgtg aaatctccaa  
 61 ctcttaacct tcaacatgaa agtctctgca gtgcttctgt gcctgctgct catgacagca  
 121 gctttcaacc cccagggact tgctcagcca gatgcactca acgtcccac tacttgctgc  
 15 181 ttcacattta gcagtaagaa gatctccttg cagaggctga agagctatgt gatcaccacc  
 241 agcaggtgtc cccagaaggc tgctcatcttc agaaccacac tgggcaagga gatctgtgct  
 301 gaccacaaagg agaagtgggt ccagaattat atgaaacacc tgggcccggaa agctcacacc  
 361 ctgaagactt gaactctgct acccctactg aaatcaagct ggagtacgtg aaatgacttt  
 421 tccattctcc tctggcctcc tcttctatgc tttggaatac ttctaccata attttcaaat  
 20 481 aggatgcatt cgggtttgtg attcaaatg tactatgtgt taagtaatat tggctattat  
 541 ttgacttggt gctggtttgg agtttatttg agtattgctg atcttttcta aagcaaggcc  
 601 ttgagcaagt aggttgctgt ctctaagccc ccttcccttc cactatgagc tgctggcagt  
 661 ggggtttgtat tcggttccca ggggttgaga gcatgcctgt gggagtcagc gacatgaagg  
 721 gatgctgcaa tgtaggaagg agagctcttt gtgaatgtga ggtgttgcta aatatgttat  
 25 781 tgtggaaaga tgaatgcaat agtaggactg ctgacatttt gcagaaaata cattttattt  
 841 aaaatctcct aaaaaaaaaa a

SEQ ID NO: 32

Amino acid sequence of human CCL13 encoded by the DNA sequence shown in SEQ ID NO: 31.

30 MKVSAVLLCLLLMTAAFN PQGLA QPDALNVPSTCCFTFSSKKISLQRLKSYVITTSRCPQ  
 KAVIFRTKL GKEICADPKEKWVQNYMKHLGRKAHTLKT

SEQ ID NO: 33

Amino acid sequence of human CCL13, a soluble active secreted form derived from SEQ ID NO: 32.

35 QPDALNVPSTCCFTFSSKKISLQRLKSYVITTSRCPQKAVIFRTKL GKEICADPKEKWVQ  
 NYMKHLGRKAHTLKT

SEQ ID NO: 34

gi|20381461|gb|BC027520.1| Mus musculus chemokine (C-C motif) ligand 12, mRNA (cDNA clone MGC:41146 IMAGE:1548072), complete cds

40 1 ttgacctcaa catgaagatt tccacacttc tatgcctcct gctcatagct accaccatca  
 61 gtcctcaggt attggctgga ccagatgcgg tgagcacccc agtcacgtgc tgttataatg  
 121 ttgtaagca gaagattcac gtccggaagc tgaagagcta caggagaatc acaagcagcc  
 181 agtgctcccc ggaagctgtg atcttcagga ccatactgga taaggagatc tgtgctgacc  
 241 ccaaggagaa gtgggttaag aattccataa accacttgga taagacgtct caaaccttca

301 tccttgaacc ttcattgtcta ggctgagagt tccaaaaact cttacgtatt tccccctgaa  
 361 gttccccacg ggcagtgtga tattttattat gatattctaaa aagagatggt ttttaataatt  
 421 taaacaaact tgcttaataa atatttaagt gtatttaagt aatatttggg ccaattaaac  
 481 cgaatctaatt ttaaaaaaaaa aaaaaaaaaa aaaaaa

## 5 SEQ ID NO: 35

Amino acid sequence of mouse CCL13 encoded by the DNA sequence shown in SEQ ID NO: 34.

MKISTLLCLLLIATTISPQVLGPDAVSTPVTCCYNVVKQKIHVRKLKSYRRITSSQCPR  
 EAVIFRTILDKEICADPKEKWVKN SINHLDKTSQTFILPSCLG

## 10 SEQ ID NO: 36

gi|27674224|ref|XM\_213425.1| Rattus norvegicus similar to small inducible cytokine A12 precursor (LOC287562), mRNA

1 ggctcctgag tcctccagct ctcattccaa agcctttggc ctcaacatga agatctccac  
 61 cttctcttgc cttctgtctca tagctgccgc catcagccca cagggtgttg ctggaccaga  
 15 121 ttcagtggtc accccagtc cctgctgtta taatgtcgt aagcagaaga tccacattcg  
 181 gaggctaaag agctacagga aaatcacaa cagccagtg ccccggaag ctgtgatctt  
 241 cagaactgta ctggataagg agctctgtgc tgaccccaag gagaagtggg ttaaggactc  
 301 catgaaccac ttggatcaga agtctcgaac tcagcaccct tgaaccttca cgtctaggct  
 361 gaaagtcca gaaaaattac atttatttcc tctgaccttc cccatggaca gtgcgatagt  
 20 421 tattttattat gatattctaaa gagagatgct ttttaataatt taaaacacaa acttactgaa  
 481 gtaatatatta atgatatcta agttatattt gggccaatta aactgacttt aattt

## SEQ ID NO: 37

Amino acid sequence of rat CCL13 encoded by the DNA sequence shown in SEQ ID NO: 36.

25 MKISTLLCLLLIAAAISPQVLGPDSVFTPTCCYNVAKQKIHRRLLKSYRKITSSQCPR  
 EAVIFRTVLDKELCADPKEKWV KDSMNHL DQKSRTQHP

## SEQ ID NO: 38

gi|22538815|ref|NM\_005623.2| Homo sapiens chemokine (C-C motif) ligand 8 (CCL8), mRNA

30 1 gtgatggaga gcaccagcaa agccttaggg cccatccctg gcctcctggt acccacagag  
 61 gggtagggcc ttggctctct tccactatga cgtcagcttc cattcttctt ttcttataga  
 121 caattttcca ttccaaggaa atcagagccc ttaatagttc agtgaggcca ctttctgtgag  
 181 cacaatccca tacccttcag cctctgctcc acagagccta agcaaaagat agaaactcac  
 241 aacttccttg ttttgttatc tggaaattat cccaggatct ggtgcttact cagcatattc  
 35 301 aaggaaggte ttacttcatt cttccttgat tgtgaccatg cccaggctct ctgctcccta  
 361 taaaaggcag gcayagccac cgaggagcag agagggtgag aacaaccag aaaccttcac  
 421 ctctcatgct gaagctcaca cccttgccct ccaagatgaa ggtttctgca gcgttctgt  
 481 gcctgctgct catggcagcc actttcagcc ctcagggact tgcctagcca gattcagttt  
 541 ccattccaat cacctgctgc tttaacgtga tcaataggaa aattcctatc cagaggctgg  
 40 601 aacggtacac aagaatcacc aacatccaat gtcccaagga agctgtgate ttcaagacca  
 661 aacggggcaa ggagggtctgt gctgacccca aggagagatg ggtcagggat tccatgaagc  
 721 atctggacca aatatttcaa aatctgaagc catgagcctt catacatgga ctgagagtca  
 781 gagcttgaag aaaagcttat ttattttccc caacctcccc cagggtgcagt gtgacattat  
 841 tttattataa catccacaaa gagattattt ttaataaatt taaagcataa tatttcttaa



901 aaagtattta attatattta agttgttgat gttttaactc tatctgtcat acatcctagt  
 961 gaatgtaaaa tgcaaaatcc tggatgatgtg ttttttgttt ttgttttcct gtgagctcaa  
 1021 ctaagttcac ggcaaaatgt cattgttctc cctcctacct gtctgtagtg ttgtgggggc  
 1081 ctcccatgga tcatcaaggt gaaacacttt ggtattcttt ggcaatcagt gctcctgtaa  
 5 1141 gtcaaatgtg tgctttgtac tgctgttggt gaaattgatg ttactgtata taactatgga  
 1201 attttgaaaa aaaatttcaa aaagaaaaaa atatataata tttaaaacta aaaaaaaaaa  
 1261 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa  
 1321 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a

SEQ ID NO: 39

- 10 Amino acid sequence of human CCL8 encoded by the DNA sequence shown in SEQ ID NO: 38.

MKVSAALLCLLLMAATFSPQGLAQPDVSIPITCCFNVINRKIPIQRLESYTRITNIQCP  
 KEAVIFKTKRGKEVCADPKERWVRDSMKHLDQIFQNLKP

SEQ ID NO: 40

- 15 Amino acid sequence of human CCL8, a soluble active secreted form derived from SEQ ID NO:39.

QPDSVSIPITCCFNVINRKIPIQRLESYTRITNIQCPKEAVIFKTKRGKEVCADPKERWV  
 RDSMKHLDQIFQNLKP

SEQ ID NO: 41

- 20 gi|20381461|gb|BC027520.1| Mus musculus chemokine (C-C motif) ligand 12, mRNA (cDNA clone MGC:41146 IMAGE:1548072), complete cds

1 ttgacctcaa catgaagatt tccacacttc tatgcctcct gctcatagct accaccatca  
 61 gtccctcagggt attggctgga ccagatgcgg tgagcaccct agtcacgtgc tgttataatg  
 121 ttgttaagca gaagattcac gtccggaagc tgaagagcta caggagaatc acaagcagcc  
 25 181 agtgtccccc ggaagctgtg atcttcagga ccatactgga taaggagatc tgtgtgaccc  
 241 ccaaggagaa gtgggttaag aattccataa accacttgga taagacgtct caaaccttca  
 301 tccttgaacc ttcattgtcta ggctgagagt tccaaaaact cttacgtatt tccccctgaa  
 361 gttccccacg ggcagtgatg tatttattat gatattctaa aagagatgtt tttataaatt  
 421 taaacaaact tgcttaataa atatttaaat gtatttaagt aatatttggg ccaattaaac  
 30 481 cgaatctaatt ttaaaaaaaaa aaaaaaaaaa aaaaaa

SEQ ID NO: 42

- Amino acid sequence of mouse CCL8 encoded by the DNA sequence shown in SEQ ID NO: 41.

35 MKISTLLCLLLIATTISPQVLGPDVSTPVTCCYNVVKQKIHVRKLKSYRRITSSQCPR  
 EAVIFRTILDKEICADPKKWKNSINHLDKTSQTFLEPSCLG

SEQ ID NO: 43

- gi|27674224|ref|XM\_213425.1| Rattus norvegicus similar to small inducible cytokine A12 precursor (LOC287562), mRNA

1 ggctcctgag tcctccagct ctcatcccaa agcctttggc ctcaacatga agatctccac

61 ccttcttttgc cttctgctca tagctgcegc catcagccca caggtgttgg ctggaccaga  
 121 ttcagtgttc accccagtc cctgctgtta taatgtcgct aagcagaaga tccacattcg  
 181 gaggctaaag agctacagga aaatcacaaag cagccagtgt ccccggaag ctgtgatctt  
 241 cagaactgta ctggataagg agctctgtgc tgacccaag gagaagtggg ttaaggactc  
 5 301 catgaaccac ttggatcaga agtctcgaac tcagatcct tgaacctca cgtctaggct  
 361 gaaagtcca gaaaaattac atttatttcc tctgaccttc cccatggaca gtgcgatagt  
 421 -tatttattat gatattctaaa gagagatgct ttaataatt taaaacacaa acttactgaa  
 481 gtaatatatta atgatattca agttatattt gggccaatta aactgacttt aattt

## 10 SEQ ID NO: 44

Amino acid sequence of rat CCL8 encoded by the DNA sequence shown in SEQ ID NO: 43.

MKISTLLCLLLIAAAISPQVLGPDVFTPTCCYNVAKQKIHRRLLKSYRKITSSQCPR  
 EAVIFRTVLDKELCADPKEKWVKDSMNHLQKSRTHP

## SEQ ID NO: 45

## 15 gi|4557017|ref|NM\_001276.1| Homo sapiens chitinase 3-like 1 (cartilage glycoprotein-39) (CHI3L1), mRNA

1 agtggagtgg gacaggtata taaaggaagt acagggcctg gggaagaggg cctgtctagg  
 61 tagctggcac caggagccgt gggcaaggga agaggccaca ccctgccctg ctctgctgca  
 121 gccagaatgg gtgtgaaggc gtctcaaaca ggctttgtgg tcctggtgct gctccagtgc  
 20 181 tgctctgcat acaaaactggt ctgctactac accagctggt cccagtaccg ggaaggcgat  
 241 gggagctgct tccagatgc ccttgaccgc ttctctgtga cccacatcat ctacagcttt  
 301 gccaatataa gcaacgatca catcgacacc tgggagtggg atgatgtgac gctctacggc  
 361 atgctcaaca cactcaagaa caggaacccc aacctgaaga ctctcttgct tgtcggagga  
 421 tggaaactttg ggtctcaaag attttccaag atagcctcca acaccagag tcgccggact  
 25 481 ttcatacaagt cagtaccgcc attcctgcgc acccatggct ttgatgggct ggaccttgcc  
 541 tggctctacc ctggacggag agacaaacag cattttacca ccctaatcaa ggaatgaag  
 601 gccgaattta taaaggaagc ccagccaggg aaaaagcagc tcctgctcag cgcagcactg  
 661 tctgctggga aggtcaccat tgacagcagc tatgacattg ccaagatata ccaacacctg  
 721 gatttcatta gcatcatgac ctacgatttt catggagcct ggcgtgggac cacaggccat  
 30 781 cacagtcccc tggtccgagg tcaggaggat gcaagtcctg acagattcag caacactgac  
 841 tatgctgtgg ggtacatggt gaggtcgggg gctcctgcca gtaagctggt gatgggcatac  
 901 cccaccttcg ggaggagctt cactctggct tcttctgaga ctggtgttgg agccccaatc  
 961 tcaggaccgg gaattccagg ccggttcacc aaggaggcag ggacccttgc ctactatgag  
 1021 atctgtgact tcctccgagg agccacagtc catagaaccc tcggccagca ggtcccttat  
 35 1081 gccaccaagg gcaaccagtg ggtaggatag gacgaccagg aaagcgtcaa aagcaagggtg  
 1141 cagtacctga aggataggca gctggcaggc gccatggtat gggccctgga cctggatgac  
 1201 ttccagggct ccttctgcgg ccaggatctg cgcttccctc tcaccaatgc catcaaggat  
 1261 gcaactcgtg caacgtagcc ctctgttctg cacacagcac gggggccaag gatgccccgt  
 1321 cccctctggt ctccagctgg ccgggagcct gatcacctgc cctgctgagt cccaggctga  
 40 1381 gcctcagtct ccctcccttg gggcctatgc agaggtccac aacacacaga tttgagctca  
 1441 gccctggtgg gcagagaggt agggatgggg ctgtggggat agtgaggcat cgcaatgtaa  
 1501 gactcgggat tagtacacac ttgttgatga ttaatggaaa tgtttacaga tccccaagcc  
 1561 tggcaaggga atttcttcaa ctccctgccc cctagccctc cttatcaaag gacaccattt  
 1621 tggcaagctc tatcaccaag gagccaaaca tcctacaaga cacagtgacc atactaatta  
 45 1681 taccctctgc aaagccagct tgaaaccttc acttaggaac gtaatcgtgt cccctatcct  
 1741 acttccccct cctaattcca cagctgctca ataaagtaca agagtttaac agtgtgttgg  
 1801 cgctttgctt tgggtctatc ttgagcgccc actagacca ctggactcac ctcccccatc  
 1861 tcttctgggt tccttctctt gagccttggg acccctgagc ttgcagagat gaaggccgcc  
 1921 atgtt

## 50 SEQ ID NO: 46

Amino acid sequence of human CHI3L1 encoded by the DNA sequence shown in SEQ ID NO: 45.

5 MGVKASQTGFVVLVLLQCCSAYKLVCCYTSWSQYREGDGSCFPDALDRFLCTHIIYSFAN  
ISNDHIDTWEWNDVTLYGMLNTLKNRNPNLKTLSSVGGWNFGSQRFSKIASNTQSRRTFI  
KSVPPFLRTHGFDGLDLAWLYPGRRDQKHFTTLIKEMKAEFIKEAQPQKKQLLLSAALSA  
GKVTIDSSYDIAKISQHLDFISIMTYDFHGAWRGTTGHHSPLFRGQEDASPDRFSNTDYA  
VGYNLRLGAPASKLVMGIPTFGRSFTLASSETGVGAPISGPGIPGRFTKEAGTLAYYEIC  
DFLRGATVHRTLGGQVPYATKGNQWVGYYDDQESVSKVQYLKDRQLAGAMVWALDLDLDFQ  
GSFCGQDLRFPLTNAIKDALAAT

10 SEQ ID NO: 47

Amino acid sequence of human CHI3L1, a soluble active secreted form derived from SEQ ID NO:46.

15 YKLVCCYTSWSQYREGDGSCFPDALDRFLCTHIIYSFANISNDHIDTWEWNDVTLYGMLN  
TLKNRNPNLKTLSSVGGWNFGSQRFSKIASNTQSRRTFIKSVPPFLRTHGFDGLDLAWLY  
PGRRDQKHFTTLIKEMKAEFIKEAQPQKKQLLLSAALSAGKVTIDSSYDIAKISQHLDFI  
SIMTYDFHGAWRGTTGHHSPLFRGQEDASPDRFSNTDYAVGYMLRLGAPASKLVMGIPTF  
GRSFTLASSETGVGAPISGPGIPGRFTKEAGTLAYYEICDFLRGATVHRTLGGQVPYATK  
GNQWVGYYDDQESVSKVQYLKDRQLAGAMVWALDLDLDFQGSFCGQDLRFPLTNAIKDALA  
AT

20 SEQ ID NO: 48

gi|33468846|ref|NM\_007695.1| Mus musculus chitinase 3-like 1 (Chi3l1), mRNA

1 cggacgcgtg ggcggacgcg tggggctggg tactaggaga agccatcatg cacacctcta  
61 ctgaagccag gatgggcatg agggcggcac tgacaggctt tgcggctcctg atgctgctcc  
121 agagctgctc tgcgtacaag ctggctctgt acttcaccag ctgggtcccag taccgggaag  
25 181 gcgttggaag cttcttaccg gacgccatcc aacctttcct gtgcaccac atcatctaca  
241 gctttgccaa catcagcagc gacaacatgc ttagcacatg ggagtggaaat gacgagtcga  
301 actatgacaa gctgaataaa ctgaagacca gaaacaccaa cctgaagacc ctctgtctg  
361 ttggaggggtg gaaatttggc gaaaaaagat tttccgagat tgcctccaac actgagagac  
421 gcactgcttt cgtccggctg gtagcccgct tcttgcgttc ttatggcttt gatgggctgg  
30 481 atctcgctg gctctaccct cgttaagag acaagcagta tttctccacc ctgatcaagg  
541 aactgaatgc ggaattcaca aaggaggctc agccaggcag agagaaactc ctgctcagc  
601 cagctttgtc agcaggaaag gtggccattg acactggcta tgacatcgcc cagatagccc  
661 aacacctgga ttttatcaat ctcatgacct acgatttcca tggagtctgg cgccaaatca  
721 caggccacca cagccccctc ttccaaggcc agaaggacac taggtttgac agatacagca  
35 781 atgtgaacta tgccgtgcag tacatgatac gtctgggagc ccaggccagc aagctactga  
841 tgggcatccc cacctttggg aagagcttca ctctggcatc ttctgaaaat cagttgggag  
901 ctccaatctc aggggaagga ttaccaggcc ggttcaccaa ggaggcaggg accctggcct  
961 actcagagat atgcgacttc ctcaaaggag ctgaagtaca tgcactctcc aacgagaagg  
1021 ttcccttcgc taccaaggcc aaccagtggg tggggtatga ggacaaggag agtgtaaaaa  
40 1081 acaagggttg gttcctgaag gagaagaagc tggcaggagc catgggtgtg gcactggatt  
1141 tggatgattt ccagggcacc tgtcagccga aggaattctt cccgctcacc aacgccatca  
1201 aggatgcctt ggcttagctc cccctttccc atatggtacc cccactctct ggccaggagt  
1261 ttaatctctt gcaatgttaa gttccccaac tgagcc!cag tttctccttc ccttggcacc  
1321 tgtgtaaggg gccacagcag gctcagctat ggagaacagg gaactagggt aggacgatgg  
45 1381 tggggttggt agagtacag tgtgagcaga tacacaaccc tggtaaggaa tgcaaattct  
1441 cagactctaa cctcccttta cccagcctga ccaaaggaca ccacttggat caagtaggca  
1501 aatatcttac aggattgagg gaccatacta attataccct ctgcaaagcc caacttgaat  
1561 ccttccctta ggaacttaat cgtcccactt ccctttccct aattccacag ctgttcaata  
1621 aagcgccaga acctaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa

## SEQ ID NO: 49

Amino acid sequence of mouse CHI3L1 encoded by the DNA sequence shown in SEQ ID NO: 48.

```

5  MGMRAALTGFAVLMLLQSCSAYKLVCFYTSWSQYREGVGSFLPDAIQPFLLCTHIYISFAN
   ISSDNMLSTWBNDESNDKLNKLTNRNTNLKTLSSVGGWKFGKRFSEIASNTERRTAF
   VRSVAPFLRSYGFGLDLAWLYPRLRDQYFSTLIKELNAEFTKEVQPGREKLLLSAALS
   AGKVAIDTGYDIAQIAHLDFINLMTYDFHGVWRQITGHHSPLFQQQKDRFDYRYSNVNY
   AVQYMIRLGAQASKLLMGIPTFGKSFTLASSENQLGAPISGEGLPGRFTKEAGTLAYYEI
10  CDFLKGAEVHRLSNEKVPFATKGNQWVGVEDKESVKNKVGFLKEKKLAGAMVWALDLDDF
   QGTCQPKFFPLTNAIKDALA

```

## SEQ ID NO: 50

gi|34880227|ref|XM\_341123.1| Rattus norvegicus chitinase 3-like 1 (cartilage glycoprotein-39) (Chi3l1), mRNA

```

15  1  gcctgaacag agggctggag ctgcagacag gagctgccgg gaatgctggg agactactgg
   61  caagaagctt tgcggtcctg atgctgtctc agagctgtct tgcgtacaaa ctggtctgtct
   121 actacaccaa ctggtccag taccgggaag gcaatgggag ctgcttccca gatgccctcg
   181 accattccct gtgcacccat atcatctaca gctttgcca catcagcaac aacaagctca
   241 gcacatcgga gtggaatgac gtaaccctgt atggcatgct gaatactctc aagaccagaa
   301 accccagact gaagacactg ctgtctgttg gaggatggag ctttggtcga gaaagatttt
20  361 ccaggattgt ctccaacgct aagagtgcga agactttcgt ccagtcggta gtcctctcc
   421 tgcggacctt tggctttgat ggactggatc tgcctggct ctaccgggc ccgaaagaca
   481 agcaacattt taccacactg atcaaggaa tgaaggcgga attcacaag gaagtcacgc
   541 caggcacaga gaaactcctg ctcaagtctg cctgttcagc aggaaaggtg acccttgaca
   601 gtggctatga tgttggccag atagcccaac acctagattt cattaatctc atgacctatg
25  661 atttccatgg aacctggcgc cacaccacag gacatcacag cccctcttc cgaggccagc
   721 aggacactgg gcctgacaga ttcagcaatg tggactatgg tgtgggttac atgctaaggc
   781 tgggagcccc caccaacaag ctagtgatgg gtatccccc ctttggaag agcttcactc
   841 tggcatcttc tgagaatcaa gtgggagctc caatcacagg gtcaggatta ccaggccgct
   901 acaccaagga gaaagggacc ctgcctact acgagatatg cgacttcctc agaggagctg
30  961 aagtacatga aattcttggc ccttgctac caagggcaac cagtgggtgg
   1021 ggatgatga cccggagagc gtcaaaaaca aggtgaagta cctgaagaac aagcagctgg
   1081 caggagccat ggtgtgggca gtggatttg atgatttccg gggctccttc tgtgggcata
   1141 acgtacactt cccgctcacc aacgccatca aggaggccct ggctgtggct tagctcccc
   1201 ttcccacat gtcccccca cctccggcc aggagctctg tcaactgcaa tgctaagtcc
35  1261 cccgactggg tgtcagtttc cctcccttgg aacctgtgta gggggccaca gcaggctcgg
   1321 acctggtgaa cagggaataa gggtaggacg gtgggattgt gaaagtcacc gtgtgagata
   1381 gatacaggtc atccctgtta atgaatgcaa attctcaagg ctcttaactc ctttccctc
   1441 atctctcccg accaaaggac accaattttg gcaagttata gcatcaagta ggcacacatc
   1501 ttacaggatt cagggaccat actaattata ccttctgcaa agcccaactt gaaaccttcc
40  1561 cttaggaact taatcgtctc tctgtccac ttcccttcc ctaattccac agctgctcaa
   1621 taaagtgccg gagcttaaca gtc

```

## SEQ ID NO: 51

Amino acid sequence of rat CHI3L1 encoded by the DNA sequence shown in SEQ ID NO: 50.

```

45  MLLQSCSAYKLVCFYTNWSQYREGNGSCFPDALDHSCLCTHIYISFANISNNKLSTSEWND
   VTLYGMLNTLKTRNPRKTLSSVGGWSFGSERFSRIVSNAKSRKTFVQSVAPFLRITYGFD
   GLDLAWLYPGPKDKQHFTTLIKELKAEFTKEVQPGTEKLLLSAAVSAGKVTLD SGYDVAQ
   IAHLDFINLMTYDFHGTWRHTTGHHSPLFRGQQDTGPD RFSNVYGVGYMLRLGAPTNK
   LVMGIPTFGKSFTLASSENQVGAPITGSGLPGRYTKBKGT LAYYEICDFLRGAEVHRI LG

```

QQVPFATKGNQWVGYYDDPESVKNKVYLKQKLAGAMVWAVDLDDFRGSFCGHNHVFPLT  
NAIKEALAVA

SEQ ID NO: 52

gi|21536275|ref|NM\_000651.3| Homo sapiens complement component (3b/4b) receptor 1,  
5 including Knops blood group system (CR1), transcript variant S, mRNA

```

1  acactctggg  cgcggagcac  aatgattggt  cactcctatt  ttcgctgagc  ttttctcttt
61  atttcagttt  tcttcgagat  caaatctggt  ttgtagatgt  gcttggggag  aatggggggc
121  tcttctccaa  gaagcccgga  gcctgtcggg  ccgccggcgc  ccggtctccc  cttctgctgc
181  ggaggatccc  tgctggcggg  tgtggtgctg  cttgcgctgc  cgggtggcctg  ggggtcaatgc
10  241  aatgccccag  aatggcttcc  atttgccagg  cctaccaacc  taactgatga  atttgagttt
301  cccattggga  catatctgaa  ctatgaatgc  cgccctgggt  attccggaag  accgttttct
361  atcatctgcc  taaaaaactc  agtctggact  ggtgctaagg  acaggtgcag  acgtaaatca
421  tgctgtaatc  ctccagatcc  tgtgaatggc  atggtgcatg  tgatcaaagg  catccagttc
481  ggatcccaaa  ttaaatattc  ttgtactaaa  ggataccgac  tcattgggtc  ctctgtctgc
15  541  acatgcatca  tctcaggtga  tactgtcatt  tgggataatg  aaacacctat  ttgtgacaga
601  attccttggt  ggctacccc  caccatcacc  aatggagatt  tcattagcac  caacagagag
661  aattttcact  atggatcagt  ggtgcaatc  cgctgcaatc  ctggaagcgg  agggagaaaag
721  gtgtttgagc  ttgtgggtga  gccctccata  tactgcacca  gcaatgacga  tcaagtgggc
781  atctggagcg  gccccgcccc  tcagtgcatt  atacctaaca  aatgcacgcc  tccaaatgtg
20  841  gaaaatggaa  tatttggtatc  tgacaacaga  agcttatttt  ccttaaata  agttgtggag
901  tttaggtgtc  agcctggctt  tgtcatgaaa  ggaccccgcc  gtgtgaagt  ccaggccctg
961  aacaaatggg  agccggagct  accaagctgc  tccagggtat  gtcagccacc  tccagatgtc
1021  ctgcagtgtg  agcgtaccca  aagggacaag  gacaactttt  cacctgggca  ggaagtgttc
1081  tacagctgtg  agcccggtta  cgacctcaga  ggggctgcgt  ctatgcgtg  cacaccccag
25  1141  ggagactgga  gccctgcagc  ccccatgt  gaagtgaat  cctgtgatga  cttcatgggc
1201  caactcttta  atggccgtgt  gctatttcca  gtaaatctcc  agcttggagc  aaaagtggat
1261  tttgtttgtg  atgaaggatt  tcaattaaaa  ggcagctctg  ctagttagct  tgtcttggct
1321  ggaatggaaa  gcctttggaa  tagcagtgtt  ccagtgtgtg  aacaaatctt  ttgtccaagt
1381  cctccagtta  ttcctaattg  gagacacaca  ggaaaacctc  tggaagtctt  tccctttgga
30  1441  aaagcagtaa  attacacatg  cgacccccag  ccagacagag  ggacgagctt  cgacctcatt
1501  ggagagagca  ccatccgctg  cacaactgac  cctcaaggga  atgggggttt  gagcagccct
1561  gcccctcgct  gtggaattct  gggctactgt  caagccccag  atcattttct  gtttgccaag
1621  ttgaaaacct  aaaccaatgc  atctgacttt  cccattggga  catctttaa  gtacgaatgc
1681  cgtcctgagt  actacgggag  gccattctct  atcacatgtc  tagataacct  ggtctggtca
35  1741  agtcccaaag  atgtctgtaa  acgtaaatca  tgtaaaactc  ctccagatcc  agtgaatggc
1801  atggtgcatg  tgatcacaga  catccagggt  ggatccagaa  tcaactattc  ttgtactaca
1861  gggcaccgac  tcattggtca  ctcatctgt  gaatgtatcc  tctcgggcaa  tgttgcccat
1921  tggagcacga  agccgccaat  ttgtcaacga  attccttgtg  ggctacccc  caccatcgcc
1981  aatggagatt  tcattagcac  caacagagag  aattttcact  atggatcagt  ggtgacctac
40  2041  cgctgcaatc  ctggaagcgg  agggagaaag  gtgtttgagc  ttgtgggtga  gccctccata
2101  tactgcacca  gcaatgacga  tcaagtgggc  atctggagcg  gccccgcccc  tcagtgcatt
2161  atacctaaca  aatgcacgcc  tccaaatgtg  gaaaatggaa  tatttggtatc  tgacaacaga
2221  agcttatttt  ccttaaata  agttgtggag  tttaggtgtc  agcctggctt  tgtcatgaaa
2281  ggaccccgcc  gtgtgaagt  ccaggccctg  aacaaatggg  agccggagct  accaagctgc
45  2341  tccaggggat  gtcagccacc  tccagatgtc  ctgcatgctg  agcgtaccca  aagggacaag
2401  gacaactttt  caccgggca  ggaagtgttc  tacagctgtg  agccgggcta  cgacctcaga
2461  ggggctgcgt  ctatgcgtg  cacaccccag  ggagactgga  gccctgcagc  ccccatgt
2521  gaagtgaat  cctgtgatga  cttcatgggc  caactcttta  atggccgtgt  gctatttcca
2581  gtaaatctcc  agcttggagc  aaaagtggat  tttgtttgtg  atgaaggatt  tcaattaaaa
50  2641  ggcagctctg  ctagttagct  tgtcttggct  ggaatggaaa  gcctttggaa  tagcagtgtt
2701  ccagtgtgtg  aacaaatctt  ttgtccaagt  cctccagtta  ttcctaattg  gagacacaca
2761  ggaaaacctc  tggaagtctt  tccctttggg  aaaacagtaa  attacacatg  cgacccccac
2821  ccagacagag  ggacgagctt  cgacctcatt  ggagagagca  ccatccgctg  cacaagtgc
2881  cctcaaggga  atgggggttt  gagcagccct  gcccctcgct  gtggaattct  gggctactgt
55  2941  caagccccag  atcattttct  gtttgccaag  ttgaaaacct  aaaccaatgc  atctgacttt
3001  cccattggga  catctttaa  gtacgaatgc  cgtcctgagt  actacgggag  gccattctct
3061  atcacatgtc  tagataacct  ggtctggtca  agtcccaaag  atgtctgtaa  acgtaaatca

```

3121 tgtaaaactc ctccagatcc agtgaatggc atggtgcatg tgatcacaga catccagggt  
 3181 ggatccagaa tcaactattc ttgtactaca gggcaccgac tcattgggtca ctcatctgct  
 3241 gaatgtatcc tctcgggcaa tgctgcccac tggagcacga agccgccaat ttgtcaacga  
 3301 attccttgtg ggctaccccc caccatcgcc aatggagatt tcattagcac caacagagag  
 5 3361 aattttcact atggatcagt ggtgacctac cgctgcaatc ctggaagcgg agggagaaag  
 3421 gtgtttgagc ttgtgggtga gccctccata tactgcacca gcaatgacga tcaagtgggc  
 3481 atctggagcg gcccgcccc tcagtgcatt atacctaaca aatgcacgcc tccaaatgtg  
 3541 gaaaatggaa tatttggtatc tgacaacaga agcttatttt ccttaaatga agttgtggag  
 3601 tttaggtgtc agcctggctt tgatcatgaaa ggaccccgcc gtgtgaagtg ccaggccctg  
 10 3661 aacaaatggg agccggagct accaagctgc tccagggtat gtcagccacc tccagatgtc  
 3721 ctgcatgtcg agcgtaccca aaggagacaag gacaactttt caccggggca ggaagtgttc  
 3781 tacagctgtg agccccgcta tgacctcaga ggggctgcgt ctatgcgctg cacaccccag  
 3841 ggagactgga gccctgcagc ccccatctgt gaagtgaat cctgtgatga cttcagtggc  
 3901 caacttctta atggcctgt gctatttcca gtaaatctcc agcttgagc aaaagtggat  
 15 3961 tttgtttgtg atgaaggatt tcaattaaaa ggcagctctg ctagtatttg tgtcttggct  
 4021 ggaatggaaa gcctttggaa tagcagtgtt ccagtgtgtg aacaaatctt ttgtccaagt  
 4081 cctccagtta ttcctaattg gagacacaca ggaaaacctc tggaggtctt tcccttggga  
 4141 aaagcagtaa attacacatg cgacccccac ccagacagag ggacgagctt cgacctcatt  
 4201 ggagagagca ccattccgctg cacaagtgc cctcaaggga atggggtttg gaggcagcct  
 20 4261 gccctcgct gtggaattct gggctactgt caagccccag atcattttct gtttgccaag  
 4321 ttgaaaaccc aaaccaatgc atctgacttt cccattggga catctttaa gtagaatgc  
 4381 cgtcctgagt actacgggag gccattctct atcacatgtc tagataacct ggtctgggtca  
 4441 agtcccaaag atgtctgtaa acgtaaatca tgtaaaactc ctccagatcc agtgaatggc  
 4501 atggtgcatg tgatcacaga catccagggt ggatccagaa tcaactattc ttgtactaca  
 25 4561 gggcaccgac tcattggtca ctcatctgct gaatgtatcc tctcaggcaa tactgcccac  
 4621 tggagcacga agccgccaat ttgtcaacga attccttgtg ggctaccccc aacctcgcc  
 4681 aatggagatt tcattagcac caacagagag aattttcact atggatcagt ggtgacctac  
 4741 cgctgcaatc ttggaagcag agggagaaag gtgtttgagc ttgtgggtga gccctccata  
 4801 tactgcacca gcaatgacga tcaagtgggc atctggagcg gccccgcccc tcagtgcatt  
 30 4861 atacctaaca aatgcacgcc tccaaatgtg gaaaatggaa tatttggtatc tgacaacaga  
 4921 agcttatttt ccttaaatga agttgtggag tttaggtgtc agcctggctt tgatcatgaaa  
 4981 ggaccccgcc gtgtgaagtg ccaggccctg aacaaatggg agccagagtt accaagctgc  
 5041 tccagggtgt gtcagccgcc tccagaaatc ctgcatggtg agcatacccc aagccatcag  
 35 5101 gacaactttt cacctgggca ggaactgttc tacagctgtg agcctggcta tgacctcaga  
 5161 ggggctgcgt ctctgcactg cacaccccag ggagactgga gccctgaagc cccgagatgt  
 5221 gcagtgaat cctgtgatga cttcttgggt caactccctc atggccgtgt gctatttcca  
 5281 cttaatctcc agcttggggc aaaggtgtcc tttgtctgtg atgaagggtt tcgcttaag  
 5341 ggcagttccg ttagtcattg tgtcttgggt ggaatgagaa gcctttggaa taacagtgtt  
 5401 cctgtgtgtg aacatatctt ttgtccaaat cctccagcta tccttaatgg gagacacaca  
 40 5461 ggaactccct ctggagatat tccctatgga aaagaaatat cttacacatg tgacccccac  
 5521 ccagagagag ggatgacctt caacctcatt ggggagagca ccattccgctg cacaagtgc  
 5581 cctcatggga atggggtttg gaggagccct gccctcgct gtgaactttc tgttcgtgct  
 5641 ggtcactgta aaaccccaga gcagtttcca tttgccagtc ctacgatccc aattaatgac  
 5701 tttgagtttc cagtcgggac atctttgaat tatgaatgcc gtcctgggta ttttgggaaa  
 45 5761 atgttctcta tctcctgcct agaaaacttg gtctggtcaa gtgttgagaa caactgtaga  
 5821 cgaaaatcat gtggacctcc accagaacct ttcaatggaa tgggtgcata aaacacagat  
 5881 acacagtttg gatcaacagt taattattct tgtaatgaag ggtttcgact cattggttcc  
 5941 ccatctacta cttgtctcgt ctcaggcatt aatgtcacat gggataagaa ggcacctatt  
 6001 tgtgagatca tatcttgtga gccacctcca accatatcca atggagactt ctacagcaac  
 50 6061 aatagaacat cttttcaca tggaacggtg gtaacttacc agtgccacac tggaccagat  
 6121 ggagaacagc tgtttgagct tgtgggagaa cgggtcaatat attgcaccag caaagatgat  
 6181 caagttgggtg tttggagcag ccctccccct cgggtgattt ctactaataa atgcacagct  
 6241 ccagaagttg aaaaatgcaat tagagtacca ggaaacagga gtttcttttc cctcactgaq  
 6301 atcgtcagat ttagatgtca gcccggggtt gtcagtgtg ggtccacac tgtgcagtgc  
 55 6361 cagaccaatg gcagatgggg gcccaagctg ccacactgct ccagggtgtg tcagccgct  
 6421 ccagaaatcc tgcatgggtga gcataaccta agccatcagg acaacttttc acctggcag  
 6481 gaagtgttct acagctgtga gccagctat gacctcagag gggctgcgtc tctgcagtgc  
 6541 acgccccagg gagactggag ccctgaagcc cctagatgta cagtgaatc ctgtgatgac  
 6601 ttctggggcc aactccctca tggcctgtgt ctacttccac ttaattctcca gcttggggca  
 60 6661 aaggtgtcct ttgtttgcga tgaaggggtc cgattaaaag gcaggtctgc tagtcattgt  
 6721 gtcttggctg gaatgaaagc cctttggaat agcagtggtc cagtgtgtga acaaatcttt

6781 tgtccaaatc ctccagctat ccttaatggg agacacacag gaactccctt tggagatatt  
 6841 ccctatggaa aagaaatata ttacgcatgc gacacccacc cagacagagg gatgaccttc  
 6901 aacctcattg gggagagctc catccgctgc acaagtgacc gtcaagggaa tgggggtttgg  
 6961 agcagccctg cccctcgctg tgaactttct gttcctgctg cctgcccaga tccaccaag  
 5 7021 atccaaaacg ggcattacat tggaggacac gtatctctat atcttcctgg gatgacaatc  
 7081 agctacattt gtgaccccggt ctacctgtta gtgggaaagg gcttcatttt ctgtacagac  
 7141 cagggaatct ggagccaatt ggatcattat tgcaaaagaa taaattgtag cttccactg  
 7201 ttatgaatg gaatctcgaa ggagttagaa atgaaaaaag tatatcata tggagattat  
 7261 gtgactttga agtgtagaaga tgggtatact ctgggaaggca gtccctggag ccagtgcag  
 10 7321 gcggatgaca gatgggaccc tcctctggcc aaatgtacct ctctgacaca tgatgctctc  
 7381 atagttggca ctttatctgg tacgatcttc tttattttac tcatcatttt cctctcttgg  
 7441 ataattctaa agcacagaaa agcaataaat gcacatgaaa accctaaaga agtggctatc  
 7501 catttacatt ctcaaggagg cagcagcggt catccccgaa ctctgcaaac aatgaagaa  
 7561 aatagcaggg tccttccttg acaaagtact atacagctga agaactctc gaatacagtt  
 15 7621 ttgggtgggaa aggagccaat tgatttcaac agaatacagat ctgagcttca taaagtcttt  
 7681 gaagtgactt cacagagacg cagacatgtg cacttgaaga tgctgcccct tccctggtag  
 7741 ctagcaaagc tcctgcctct ttgtgtgcgt cactgtgaaa cccccaccct tctgcctcgt  
 7801 gctaaacgca cacagtatct agtcagggga aaagactgca tttaggagat agaaaaatag  
 7861 ttggattact taaagggaata aggtgttgcc tggaaattct gggttgtaag gtggtcactg  
 20 7921 tcttttttta aaatatttgt aatatggaat gggctcagta agaagagctt ggaaaatgca  
 7981 gaaagttagt aaaaaataagt cacttataat tatgctacct actgataacc actcctaata  
 8041 ttttgattca tttctgcct atcttcttcc acatatgtgt tttttacat acgtcctttt  
 8101 cccccagtt tgtttccttt tattttatag agcagaacct tagtctttta aaccagttta  
 8161 gagtgaata tatgctatat cagtttttac tttctctagg gagaaaaatt aatttactag  
 25 8221 aaaggcatga aatgatcatg ggaagagtgg ttaagactac tgaagagaaa tatttggaaa  
 8281 ataagatttc gatattctct ttttttttga gatggagtct ggctctgtct cccaggtgg  
 8341 agtgcagtgg cgtaatctcg gctcactgca agctccgct cctgggttga caccatttcc  
 8401 ctgcctcagc ctccctgagta gttgggacta ccagtagatg ggactacagg cacctgcca  
 8461 cagccccggc taattttttt gtatttttag tagagacggg gtttcacat gtagccagg  
 30 8521 atgggtctgga tctcctgacc tcgtgatcca cccgctcgg cctcccaaag tgctgcgatt  
 8581 acaggcatga gccaccgagc ctggccgctt tcgatatttt ctaaacttta attcaaaagc  
 8641 actttgtgct gtgttctata taaaaaacat aataaaaatt gaaatgaaag aataattgtt  
 8701 attataaaag tactagctta ctttctgatg gattcagaat atactaaatt aacttttaa  
 8761 aacacaactt ttaaaaaatg tatcaaaaaa aataaacgtg ttctgatatt tta

SEQ ID NO: 53

Amino acid sequence of human CR1 encoded by the DNA sequence shown in SEQ ID NO: 52.

40 MGASSPRSEPPVGGPPAPGLPFCGGSL LAVVLLALPVAWGQCNAPEWL PPARPTNL TDE  
 FEPPIGTYL NYECRPGYSGRPFSII CLKNSVWTGAKDRCKR KSCRNPPDPVNGMVHVIK  
 IQFGS QIKYSCTKG YRLIGSSSATCII SGGTVIWDNETPICDRIPCGLPPTITNGDFIST  
 NREN FHYGSVV TYRCNPGSGGRKV FELVGEPSIYCTSNDDQVGIWSGPAPQCIIPNKCTP  
 PNVEN GILVSDN RSLFSLNEVVEFR CQPGFVMKGPRRVK CQALNKWEPELPSCSRVCQPP  
 45 PDVLHAERTQRDKDNFSPGQEVFYSC EPGYDLRGAASMRCTPQGDWSPAAPTCEVKS CDD  
 FMGQLLNGRVLFPVNLQ LGAKVDFVCDEGFQLKGSSASYCVLAGMESLWNSSVPVCEQIF  
 CPSP PVI PNGRHTGKPLEVFPFGKAVNYTC DPHPDRTSFDLIGESTIRCTSDPQNGVW  
 SSPAPRCGILGHCQAPDHFLFAKLKTQTNASDFPIG TSLKYECRPEYYGRPFSITCLDNL  
 VWSSPKDVCKRKSKCTPPDPVNGMVHVIDIQVGS RINYSCTTGHR LIGHSSAECILSGN  
 50 AAHWSTKPPICQRI PCGLPPTIANGDFISTNREN FHYGSVV TYRCNPGSGGRKV FELVGE  
 PSIYCTSNDDQVGIWSGPAPQCIIPNKCTPPNVEN GILVSDN RSLFSLNEVVEFR CQPGF  
 VMKGPRRVK CQALNKWEPELPSCSRVCQPPDV LHAERTQRDKDNFSPGQEVFYSC EPGY  
 55 DLRGAAASMRCTPQGDWSPAAPTCEVKS CDDFMGQLLNGRVLFPVNLQ LGAKVDFVCDEGF  
 QLKGSSASYCVLAGMESLWNSSVPVCEQIFCPSP PVI PNGRHTGKPLEVFPFGKAVNYTC  
 DPHPDRTSFDLIGESTIRCTSDPQNGVWSSPAPRCGILGHCQAPDHFLFAKLKTQTNA  
 SDFPIG TSLKYECRPEYYGRPFSITCLDNLVWSSPKDVCKRKSKCTPPDPVNGMVHVIDI  
 IQVGS RINYSCTTGHR LIGHSSAECILSGNAAHWSTKPPICQRI PCGLPPTIANGDFIST  
 NREN FHYGSVV TYRCNPGSGGRKV FELVGEPSIYCTSNDDQVGIWSGPAPQCIIPNKCTP

PNVENGILVSDNRSLSLNEVVEFRCQPGFVMKGP RRVKCQALNKWPELPSCSRVCQPP  
 PDVLHAERTQRDKDNFSPGQEVFYSCEPGYDLRGAASMRCTPQGDWSPAAPTCEVKSCDD  
 FMGQLLNGRVLFVNLQLGAKVDFVCEGFLKGSSASYCVLAGMESLWNSSVPVCEQIF  
 5 CSPSPVPIPNGRHTGKPLEVFFPGKAVNYTCDPHDRGTSFDLIGESTIRCTSDPQNGVW  
 SSPAPRCGILGHCOAPDHFLEFAKLKTQTNASDFPIGTSKYECRPEYYGRPFSITCLDNL  
 VWSSPKDVCKRKSKCTPPDPVNGMVHVITDIQVGSRIYNSCTTGHRLLIGHSSAECILSGN  
 TAHWSTKPPICQRI PCGLPPTIANGDFISTNRENFHYGSVVTYRCNLGSRGRKVPFELVGE  
 PSYICTSNDDQVGIWSGPAPQCIIPNKCTPPNVENGILVSDNRSLSLNEVVEFRCQPGF  
 10 VMKGP RRVKCQALNKWPELPSCSRVCQPPPEILHGEHTPSHODNFSPGQEVFYSCEPGY  
 DLRGAASLHCTPQGDWSPAAPRCVAVKSCDDFLGQLPHGRVLFPLNLQLGAKVSFVCEGFL  
 RLKGSVSVSHCVLVGMRLWNSSVPVCEHIFCPNPPAILNGRHTGTSPGDI PYGKEISYTC  
 DPHDRGMTFNLIGESTIRCTSDPHNGVWSSPAPRCESVRAGHCKTPEQFPFASPTIP  
 INDPEFPVGTSLNYECRPGYFGKMFSISCLENLVWSSVEDNCRRKSCGPPPEPFNGMVHI  
 15 NTDQTQGSTVNYSCNEGRLIGSPSTTCLVSGNNVTWDDKAPICEIISCEPPPTISNGDF  
 YSNRSTFHN GTVVTYQCHTGP DGEQLFELVGERSIYCTSKDDQVGVWSSPPPRCISTNK  
 CTAPEVENAIRVPGNRSFFSLTEIVRFRCPGFMVVGSHTVQCQTNGRWGPKLPHCSRVC  
 QPPPEILHGEHTLSHODNFSPGQEVFYSCEPSYDLRGAASLHCTPQGDWSPAAPRCTVKS  
 CDDFLGQLPHGRVLLPLNLQLGAKVSFVCEGFLKGRSASHCVLAGMKALWNSSVPVCE  
 20 QIFCPNPPAILNGRHTGTTPFGDI PYGKEISYACDTHPDRGMTFNLIGESSIRCTSDRQGN  
 GVWSSPAPRCESVRPAACPDPPKIQNGHYIGGHVSLYLPGMTISYICDPGYLLVGKGFIF  
 CTDQGIWSQLDHYCKEVNCSFPLFMNGISKELEMKKVYHYGDYVTLKCEDGYTLEGSPWS  
 QCQADDRWDPLAKCTSAHDALIVGTLSGTIFILLIIFLSWIILKHKRGNNAHENPKE  
 VAIHLHSQGGSSVHPRTLQTNEENSRVLP

SEQ ID NO: 54

25 gi|18490996|ref|NM\_000573.2| Homo sapiens complement component (3b/4b) receptor 1,  
 including Knops blood group system (CR1), transcript variant F, mRNA

1 acactctggg cgccggagcac aatgattggt cactcctatt ttcgctgagc ttttcctctt  
 61 atttcagatt tcttcagat caaatctggt ttgtagatgt gcttggggag aatggggggc  
 121 tcttctccaa gaagcccggg gcctgtcggg ccgcccgcgc ccggtctccc cttctgctgc  
 30 181 ggaggatccc tgctggcggg tgtggtgctg cttgcgctgc cgggtggcctg ggttcaatgc  
 241 aatgccccag aatggcttcc atttgccagg cctaccaacc taactgatga gtttgagttt  
 301 cccattggga catatctgaa ctatgaatgc cgccctggtt attccggaag accgttttct  
 361 atcatctgcc taaaaaactc agtctggact ggtgctaagg acagggtgcag acgtaaatca  
 421 tgtcgtaatc ctccagatcc tgtgaatggc atggtgcatg tgatcaaagg catccagttc  
 35 481 ggatcccaaa ttaaatattc ttgtactaaa ggataccgac tcattgggtc ctgctctgcc  
 541 acatgatca tctcaggtga tactgtcatt tgggataatg aaacacctat ttgtgacaga  
 601 attccttgtg ggctaccccc caccatcacc aatggagatt tcattagcac caacagagag  
 661 aattttcact atggatcagt ggtgacctac cgctgcaatc ctggaagcgg agggagaaag  
 721 gtgtttgagc ttgtgggtga gccctccata tactgcacca gcaatgacga tcaagtgggc  
 40 781 atctggagcg gccccgccc tcagtgcatt atacctaaca aatgcacgcc tccaaatgtg  
 841 gaaaatggaa tattggtatc tgacaacaga agcttatttt ccttaaatga agttgtggag  
 901 tttaggtgtc agcctggctt tgtcatgaaa ggaccccgcc gtgtgaagtg ccaggccctg  
 961 aacaaatggg agccggagct accaagctgc tccagggtat gtcagccacc tccagatgtc  
 1021 ctgcatgctg agcgtaccca aagggacaag gacaactttt cacctgggca ggaagtgttc  
 45 1081 tacagctgtg agcccggcta cgacctcaga ggggctgctg ctatgctgtg cacaccccag  
 1141 ggagactgga gccctgcagc ccccatatgt gaagtgaat cctgtgatga cttcatgggc  
 1201 caacttctta atggccgtgt gctatttcca gtaaatctcc agcttgagc aaaagtggat  
 1261 tttgtttgtg atgaaggatt tcaattaaaa ggcagctctg ctagttactg tgtcttggct  
 1321 ggaatggaaa gcctttggaa tagcagtgtt ccagtgtgtg aacaaatctt ttgtccaagt  
 50 1381 cctccagtta ttcctaattg gagacacaca ggaaaacctc tggaagtctt tccctttgga  
 1441 aaagcagtaa attacacatg cgacccccac ccagacagag ggacgagctt cgacctcatt  
 1501 ggagagagca ccattccgtg cacaagtgc cctcaaggga atgggggttg gagcagccct  
 1561 gccctcgct gtggaattct gggctcactg caagccccag atcattttct gtttgccaag  
 1621 ttgaaaaccc aaaccaatgc atctgacttt cccattggga catcttttaa gtacgaatgc  
 55 1681 cgtcctgagt actacgggag gccattctct atcacatgtc tagataacct ggtctggtca  
 1741 agtcccaaaag atgtctgtaa acgtaaatca tgtaaaactc ctccagatcc agtgaatggc  
 1801 atggtgcatg tgatcacaga catccaggtt ggatccagaa tcaactattc ttgtactaca



	1861	gggcaccgac	tcattggtca	ctcatctgct	gaatgtatcc	tctcgggcaa	tgctgcccat
	1921	tggagcacga	agccgccaat	ttgtcaacga	attccttggtg	ggctaccccc	caccatcgcc
	1981	aatggagatt	tcattagcac	caacagagag	aatttttact	atggatcagt	ggtgacctac
	2041	cgctgcaatc	ctggaagcgg	agggagaaaag	gtgtttgagc	ttgtgggtga	gccctccata
5	2101	tactgcacca	gcaatgacga	tcaagtgggc	atctggagcg	gcccggcccc	tcagtgcatt
	2161	atacctaaca	aatgcacgcc	tccaaatgtg	gaaaatggaa	tatttggtatc	tgacaacaga
	2221	agcttattttt	ccttaaataga	agttgtggag	tttaggtgtc	agcctggcct	tgtcatgaaa
	2281	ggacccccgcc	gtgtgaagtg	ccaggccctg	aacaaatggg	agccggagct	accaagctgc
	2341	tccagggtat	gtcagccacc	tccagatgtc	ctgcatgctg	agcgtaccca	aaggggacaag
10	2401	gacaactttt	caccgggga	ggaagtgttc	tacagctgtg	agcccggtc	tgacctcaga
	2461	cgggctgcgt	ctatgcgctg	cacaccccg	ggagactgga	gccctgcagc	ccccacatgt
	2521	gaagtgaat	cctgtgatga	cttcatgggc	caacttctta	atggcctgtg	gctatttcca
	2581	gtaaatctcc	agcttgagc	aaaagtggat	tttgtttgtg	atgaaggatt	tcaattaaaa
	2641	ggcagctctg	ctagttattg	tgtcttggtc	ggaatggaaa	gcctttggaa	tagcagtgtt
15	2701	ccagtgtgtg	aacaaatcct	ttgtccaaag	cctccagtta	ttcctaattg	gagacacaca
	2761	ggaaaacctc	tggaagtctt	tccctttgga	aaagcagtaa	attacacatg	cgacccccac
	2821	ccagacagag	ggacgagctt	cgacctcatt	ggagagagca	ccatccgctg	cacaagtgc
	2881	ctccaaggga	atgggggttg	gagcagccct	gcccctcgct	gtggaattct	gggtcactgt
	2941	caagccccag	atcattttct	gtttgccaa	ttgaaaaccc	aaaccaatgc	atctgacttt
20	3001	cccattggga	catctttaaa	gtacgaatgc	cgctctgagt	actacgggag	gccatttctt
	3061	atcacatgtc	tagataacct	ggtctggtca	agtcccaaag	atgtctgtaa	acgtaaatca
	3121	tgtaaaatc	ctccagatcc	agtgaatggc	atgggtgcag	tgatcacaga	catccaggtt
	3181	ggatccagaa	tcaactattc	ttgtactaca	gggcaccgac	tcattggtca	ctcatctgct
	3241	gaatgtatcc	tctcaggcaa	tactgcccac	tggagcacga	agccgccaat	ttgtcaacga
25	3301	attccttggtg	ggctaccccc	aaccatcgcc	aatggagatt	tcattagcac	caacagagag
	3361	aatttttact	atggatcagt	ggtgacctac	cgctgcaatc	ttggaagcag	agggagaaaag
	3421	gtgtttgagc	ttgtgggtga	gccctccata	tactgcacca	gcaatgacga	tcaagtgggc
	3481	atctggagcg	gccccgcccc	tcagtgcatt	atacctaaca	aatgcacgcc	tccaaatgtg
	3541	gaaaatggaa	tatttggtatc	tgacaacaga	agcttatttt	ccttaaataga	agttgtggag
30	3601	tttaggtgtc	agcctggcct	tgtcatgaaa	ggacccccgc	gtgtgaagtg	ccaggccctg
	3661	aacaaatggg	agccagagtt	accaagctgc	tccagggtgt	gtcagccgcc	tccagaaatc
	3721	ctcgactgtg	agcatacccc	aagccatcag	gacaactttt	cacctgggca	ggaagtgttc
	3781	tacagctgtg	agcctggcta	tgacctcaga	ggggctgcgt	ctctgcactg	cacagcccag
	3841	ggagactgga	gccctgaagc	cccagatgtg	gcagtgaat	cctgtgatga	cttcttgggt
35	3901	caactccctc	atggccgtgt	gctatttcca	cttaatctcc	agcttggggc	aaaggtgtcc
	3961	tttgtctgtg	atgaagggtt	tcgcttaaag	ggcagttccg	ttagtcattg	tgtcttgggt
	4021	ggaatgagaa	gcctttggaa	taacagtgtt	cctgtgtgtg	aacatatctt	ttgtccaaat
	4081	cctccagcta	tccttaatgg	gagacacaca	ggaactccct	ctggagatat	tccctatgga
	4141	aaagaaaatat	cttacacatg	tgacccccac	ccagacagag	ggatgacctt	caacctcatt
40	4201	ggggagagca	ccatccgctg	cacaagtgc	cctcatggga	atgggggttg	gagcagccct
	4261	gcccctcgct	gtgaactttc	tgctcgtgct	ggtcactgta	aaaccccaga	gcagtttcca
	4321	tttgccagtc	ctacgatccc	aattaatgac	tttgagtgtc	cagtcgggac	atctttgaa
	4381	tatgaatgcc	gtcctgggta	ttttgggaaa	atgttctcta	tctcctgcct	agaaaacttg
	4441	gtctggtcaa	gtgttgaaga	caactgtaga	cgaaaatcat	gtggacctcc	accagaaccc
45	4501	ttcaatggaa	tggtgcatat	aaacacagat	acacagtgtt	gatcaacagt	taattattct
	4561	tttaagtgaag	ggtttcgact	cattgggtcc	ccatctacta	cttgtctcgt	ctcaggcaat
	4621	aatgtcacat	gggataagaa	ggcacttatt	tgtgagatca	tatcttgtga	gccacctcca
	4681	accatatcca	atggagactt	ctacagcaac	aatagaacat	cttttcacaa	tggaacggtg
	4741	gtaacttacc	agtgccacac	tggaccagat	ggagaacagc	tgtttgagct	tggtgggagaa
50	4801	cgggtcaatat	attgcaccag	caaagatgat	caagtgggtg	tttggagcag	ccctccccct
	4861	cgggtgatatt	ctactaataa	atgcacagct	ccagaagtgt	aaaatgcaat	tagagtacca
	4921	ggaaacagga	gtttcttttc	cctcactgag	atcatcagat	ttagatgtca	gcccgggttt
	4981	gtcatggtag	ggccccacac	tgctgagctg	cagaccaatg	gcagatgggg	gcccagctg
	5041	ccacactgct	ccagggtgtg	tcagccgctc	ccagaaatcc	tgcatgggtga	gcatacccta
55	5101	agccatcagg	acaacttttc	acctgggcag	gaagtgttct	acagctgtga	gcccagctat
	5161	gacctcagag	gggctgcgtc	tctgcactgc	acgccccagg	gagactggag	ccctgaagcc
	5221	cctagatgta	cagtgaatac	ctgtgatgac	ttcctgggcc	aactccctca	tgccgctgtg
	5281	ctacttccac	ttaatctcca	gcttggggca	aaggtgtcct	ttgtttgcga	tgaagggttc
	5341	cgattaaaag	gcaggtctgc	tagtcattgt	gtcttggctg	gaatgaaagc	cctttggaat
60	5401	agcagtgttc	cagtgtgtga	acaaatcttt	tgtccaaatc	ctccagctat	ccttaatggg
	5461	agacacacag	gaactccctt	tggagatatt	ccctatggaa	aagaaatate	ttacgcatgc

5521 gacacccacc cagacagagg gatgacctc aacctcattg gggagagctc catccgctgc  
 5581 acaagtgacc ctcaagggaa tgggggttgg agcagccctg cccctcgctg tgaactttct  
 5641 gttcctgctg cctgccaca tccaccaag atccaaaacg ggcattacat tggaggacac  
 5701 gtatctctat atcttcttgg gatgacaatc agctacactt gtgaccccg gtacctgtta  
 5761 gtgggaagg gcttcatttt ctgtacagac cagggaatct ggagccaatt ggatcattat  
 5821 tgcaaaagaag taaattgtag cttccactg tttatgaatg gaatctcgaa ggagtttaga  
 5881 atgaaaaaag tatatcacta tggagattat gtgactttga agtgtgaaga tgggtatact  
 5941 ctggaaggca gtccctggag ccagtgccag gcggatgaca gatgggacct tcctctggcc  
 6001 aaatgtacct ctctgcaca tgatgtcttc atagtggca ctttatctgg tacgatcttc  
 6061 tttattttac tcatcatttt cctctcttgg ataattctaa agcacagaaa aggcaataat  
 6121 gcacatgaaa accctaaaga agtggctatc catttacatt ctcaaggagg cagcagcgtt  
 6181 catccccgaa ctctgcaaac aaatgaagaa aatagcaggg tccttccttg acaagttact  
 6241 atacagctga agaactctc gaatacaatt ttgggtggaa aggagccaat tgatttcaac  
 6301 agaatcagat ctgagcttca taaagtcttt gaagtgactt cacagagacg cagacatgtg  
 6361 cacttgaaaga tgctgcccct tccctggtag ctagcaaaag tcctgcctct ttgtgtgctg  
 6421 cactgtgaaa cccccacct tctgcctcgt gctaaacgca cacagtatct agtcagggga  
 6481 aaagactgca tttaggagat agaaaatagt ttggattact taaaggaata aggtgttgcc  
 6541 tgggaatttct ggtttgtgaa gtggtcactg ttctttttta aaatatttgt aatatggaat  
 6601 gggctcagta agaagagctt ggaaaatgca gaaagttagt aaaaataagt cacttataat  
 6661 tatgctacct actgataacc actcctaata ttttgattca ttttctgcct atcttcttct  
 6721 acatatgtgt ttttttacct acgtactttt ccccccctag tttgttttct tttattttat  
 6781 agagcagaac cctagtcttt taaacagttt agagtgaat atatgctata tcagttttta  
 6841 ctttctctag ggagaaaaat taatttacta gaaaggcatg aaatgatcat gggaagagtg  
 6901 gtaagacta ctgaagagaa atatttggaa aataagattt cgatatcttc ttttttttg  
 6961 agatggagtc tggctctgtc tcccaggctg gagtgcagtg gcgtaatctc ggctcactgc  
 7021 aacgtccgcc tcccgggttg acaccatttt cctgcctcag cctcctgagt agttgggact  
 7081 accagtagat gggactacag gcacctgcca acacgcccg ctaatttttt tgtattttta  
 7141 gttagagcgg ggtttcacca tgttagccag gatggtctgg atctcctgac cctgtgtacc  
 7201 accgcctcgg gcctcccaa gtgctgcatg tacaggcatg agccaccgag cctggccgct  
 7261 ttgatatttt tctaaacttt aattcaaaag cactttgtgc tgtgttctat ataaaaaaca  
 7321 taataaaaaat tgaatgaaa gaataattgt tattataaaa gtactagctt acttttgtat  
 7381 ggattcagaa tatactaaat taacttttta aaacacaact ttaaaaaaat gtatcaaaaa  
 7441 taataaacgt gttctgatat tttta

SEQ ID NO: 55

35 Amino acid sequence of human CR1 variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 54.

40 MGASSPRSPPEVPGPPAPGLPFCGGSLLAVVLLALPVANGQCNAPWLPPFARPTNLTDE  
 FEFPIGTYLNYECRPGYSGRPFSIIICLNKSVWTGAKDRCKRKSERNPPDPVNGMVHVIK  
 IQFGSQIKYSCTKGYRLIGSSSATCIISGDTVIWNETPICDRIPGLPPTITNGDFIST  
 NRENPHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIIWSGPAPQCIIPNKCTP  
 PNVENGILVSDNRSLSLNEVVEFRCPQPGFVMKGPRRVKQALNKWEPELPSCSRVCQPP  
 PDVLHAERTQRDKDNFSPGQEVFYSCEPGYDLRGAASMRCTPQGDWSPAAPTCEVKSCDD  
 FMGQLLNGRVLFVNLQLGAKVDFVCEGFLKGSSASYCVLAGMESLWNSSVPVCEQIF  
 45 CPSPPIVINGRHTGKPLEVFPFGKAVNYTCDFPDRGTSFDLIGESTIRCTSDPQNGVW  
 SSPAPRCGILGHCAQPDHFLPAKLKTQTNASDFPIGTSKYECRPEYYGRPFSITCLDNL  
 VWSSPKDVCKRKSCKTPDPVNGMVHVIDIQVGSRIYNSCTTGRLIGHSSAECILSGN  
 AAHWSTKPPICQRIIPGLPPTIANGDFISTNRENPHYGSVVTYRCNPGSGGRKVFELVGE  
 PSYCTSNDDQVGIIWSGPAPQCIIPNKCTPPNVENGILVSDNRSLSLNEVVEFRCPQPGF  
 VMKGPRRVKQALNKWEPELPSCSRVCQPPPDVLHAERTQRDKDNFSPGQEVFYSCEPGY  
 50 DLRGAASMRCTPQGDWSPAAPTCEVKSCDDFMGQLLNGRVLFVNLQLGAKVDFVCEGFL  
 QKGSSASYCVLAGMESLWNSSVPVCEQIFCPSPPIVINGRHTGKPLEVFPFGKAVNYTC  
 DPHDRGTSFDLIGESTIRCTSDPQNGVWSSPAPRCGILGHCAQPDHFLPAKLKTQTN  
 SDFPIGTSKYECRPEYYGRPFSITCLDNLVWSSPKDVCKRKSCKTPDPVNGMVHVIDI  
 IQVGSRIYNSCTTGRLIGHSSAECILSGNTAHWSTKPPICQRIIPGLPPTIANGDFIST  
 55 NRENPHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVGIIWSGPAPQCIIPNKCTP  
 PNVENGILVSDNRSLSLNEVVEFRCPQPGFVMKGPRRVKQALNKWEPELPSCSRVCQPP  
 PEILHGEHTPSHQDNFSPGQEVFYSCEPGYDLRGAASLHCTPQGDWSPAAPCAVKSCDD

FLGQLPHGRVLPPLNLQLGAKVSFVCDGFRLLKGSSVSHCVLVGMRLWNNSVPVCEHIF  
 CPNPPAILNGRHTGTSPGDIPIYGKEISYTCDPHPDRGMTFNLIGESTIRCTSDPHNGVW  
 SSPAPRCELSVRAGHCKTPEQFPFASPTIPINDFEFVGTSLNYECRPGYFGKMFSISCL  
 5 ENLVWSSVEDNCRKSCGPPPEPFNGMVHINTDTQFGSTVNYSCNEGFRLLIGSPSTTCLV  
 SGNNVTWDKKAPICEIISCEPPPTISNGDFYSNNRTSFHNGTVVITYQCHTGPDGEQLFEL  
 VGERSIYCTSKDDQGVWSSPPPRCISTNKCTAPEVENAIRVPGNRSFSLTEIIRFCQ  
 PGFVMVGSHTVQCQTNGRWGPKLPHCSRVCQPPPEILHGEHTLSHQDNFSPGOEVFYSC  
 PSYDLRGAASLHCTPQGDWSPEAPRCTVKSCDDFLGQLPHGRVLLPLNLQLGAKVSFVCD  
 10 EGFRLKGRSASHCVLAGMKALWNSSVPVCEQIFCPNPPAILNGRHTGTSPGDIPIYGKEIS  
 YACDTHPDRGMTFNLIGESSIRCTSDPQNGVWSSPAPRCELSVPAACPHPPKIQNGHYI  
 GGHVSLYLPGMTISYTCDPGYLLVGKGFIFCTDQGIWSQLDHYCKEVCNCSFPLFMNGISK  
 BLEMKKVYHYGDYVTLKCEDGYTLGSPWSQCQADDRWDPPLAKCTSRADALIVGTLG  
 TIFPILLIIFLSWIILKHKRGNNAHENPKEVAIHLHSQGGSSVHPRTLQTNEENSRLVLP

SEQ ID NO: 56

15 gi|27262658|ref|NM\_005211.2| Homo sapiens colony stimulating factor 1 receptor, formerly  
 McDonough feline sarcoma viral (v-fms) oncogene homolog (CSF1R), mRNA

1 gaagggcaga cagagtgtcc aaaagcgtga gagcacgaag tgaggagaag gtggagaaga  
 61 gagaagagga agaggaagag gaagagagga agcggaggga actgcggcca ggctaaaagg  
 121 ggaagaagag gatcagccca aggaggagga agaggaaaac aagacaaaaca gccagtgcag  
 20 181 aggagaggaa cgtgtgtcca gtgtcccgat ccctgcggag ctagtagctg agagctctgt  
 241 gccctgggca ccttgagcc ctgcacctgc ctgccacttc cccaccgagg ccatgggccc  
 301 aggagtctctg ctgctcctgc tgggtggccac agcttggcat ggtagggaa tccagtgat  
 361 agagcccgagt gtccctgagc tgggtcgtgaa gccaggagca acggtgacct tgcgatgtgt  
 421 gggcaatggc agcgtggaat gggatggccc cccatcacct cactggaccc tgtactctga  
 25 481 tggctccagc agcatcctca gcaccaaca cgctaccttc caaaacacgg ggacctatcg  
 541 ctgcaactgag cctggagacc ccttggagg cagcgccgccc atccacctct atgtcaaaga  
 601 ccttggccgg ccctggaacg tgctagcaca ggagggtggc gtgttcgagg accaggagc  
 661 actactgccc tgtctgctca cagaccgggt gctggaagca ggcgtctcgc tgggtgcgtgt  
 721 gcgtggccgg cccctcatgc gccacaccaa ctactccttc tcgccctggc atggcttcac  
 30 781 catccacagg gccaatgcca ttcagagcca ggactatcaa tgtagtgccc tgatgggtgg  
 841 caggaagggtg atgtccatca gcatccggct gaaagtgcag aaagtcatcc cagggccccc  
 901 agccttgaca ctggtgcctg cagagctggt gcggattcga ggggaggctg cccagatcgt  
 961 gtgctcagcc agcagcgttg atgttaactt tgatgtcttc ctccaaaca acaacaccaa  
 1021 gctcgcaatc cctcaacaat ctgactttca taataaccgt taccaaaaag tcctgacctt  
 35 1081 caacctcgat caagtagatt tccaacatgc cggcaactac tcctgcgtgg ccagcaacgt  
 1141 gcagggcaag cactccacct ccatgttctt ccgggtggta gagagtgcct acttgaactt  
 1201 gagctctgag cagaacctca tccaggagggt gaccgtgggg gaggggctca acctcaaagt  
 1261 catgggtggag gcctaccag gcctgcaagg ttttaactgg acctacctgg gaccttttc  
 1321 tgaccaccag cctgagccca agcttgctaa tgctaccacc aaggacacat acaggcacac  
 40 1381 cttcaccctc tctctgcccc gcctgaagcc ctctgaggct ggccgctact ccttctggc  
 1441 cagaaaccca ggaggctgga gagctctgac gtttgagctc acccttcgat acccccaga  
 1501 ggtaagcgtc atatggacat tcatcaacgg ctctggcacc cttttgtgtg ctgcctctgg  
 1561 gtacccccag cccaacgtga catggctgca gtgcagtggc cactactgata ggtgtgatga  
 1621 ggcccaagtg ctgcaggctt gggatgaccc atacctgag gtctgagcc aggagccctt  
 45 1681 ccacaagggtg acggtgcaga gcctgctgac tggtagagac ttagagcaca accaaacct  
 1741 cgagtgcagg gcccaacaac gcgtggggag tggctcctgg gccttcatac ccatctctgc  
 1801 aggagccac acgcatcccc cggatgagtt cctcttcaca ccagtgggtg tgcctgcct  
 1861 gtccatcatg gccttgctgc cctgctgctc cctgctgcta ttgtacaagt ataagcagaa  
 1921 gcccaggtac cagggtccgt ggaagatcat cgagagctat gagggcaaca gttatacttt  
 50 1981 catcgacccc acgagctgc cttacaacga gaagtgggag tcccccgga acaacctgca  
 2041 gtttggtgtaag acctcggag ctggagcctt tgggaagggt gtggaggcca cggcctttgg  
 2101 tctgggcaag gaggatgctg tcctgaagggt ggctgtgaag atgctgaagt ccacggccca  
 2161 tgctgatgag aaggaggccc tcatgtccga gctgaagatc atgagccacc tgggcccaga  
 2221 cgagaacatc gtcaaccttc tgggagcctg taccatgga ggccctgtac tggcatcac  
 55 2281 ggagtactgt tgctatggcg acctgctcaa ctttctgcca aggaaggctg aggccatgct  
 2341 gggacccagc ctgagccccc gccaggaccc cgaggagggc gtcgactata agaactcca  
 2401 cctcgagaag aaatatgtcc gcaggagacag tggcttctcc agccagggtg tggacaccta

```

2461 tgtggagatg aggcctgtct ccacttcttc aaatgactcc ttctctgagc aagacctgga
2521 caaggaggat ggacggcccc tggagctccg ggacctgctt cacttctcca gccaaagtagc
2581 ccagggcacg gccttcctcg cttccaagaa ttgcatccac cgggacgtgg cagcgcgtaa
5 2641 cgtgctgttg accaatggtc atgtggccaa gattggggac ttggggctgg ctaggggacat
2701 catgaatgac tccaactaca ttgtcaaggg caatgcccgc ctgcctgtga agtggatggc
2761 cccagagagc atctttgact gtgtctacac ggttcagagc gacgtctggt cctatggcat
2821 cctcctctgg gagatcttct cacttgggct gaatccctac cctggcatcc tgggtgaacag
2881 caagttctat aaactgggtg aggatggata ccaaattggcc cagcctgcat ttgccccaaa
2941 gaatatatac agcatcatgc aggcctgctg ggcccttgag cccaccacac gaccacacct
10 3001 ccagcagatc tgcctcttcc ttcaggagca ggcccaagag gacaggagag agcgggacta
3061 taccaatctg ccgagcagca gcagaagcgg tggcagcggc agcagcagca gtgagctgga
3121 ggaggagagc tctagttagc acctgacctg ctgagagcaa ggggatatcg cccagccctt
3181 gctgcagccc aacaactatc agttctgtcg aggagttagc gacaggaggt accactctcc
3241 cctcctccaa acttcaactc cttcatggat ggggagacac ggggagaaca tacaactctt
15 3301 gccttcggtc atttcaactc acagctcgcc ccagctctga aacttgggaa ggtgagggat
3361 tcaggggagg tcagaggatc ccacttctcg agcatgggccc atcactgcca gtcaggggct
3421 gggggctgag ccctcaccct cccctccctt actgttctca tgggtgttggc ctctgttttg
3481 ctatgccaac tagtagaacc ttcttctcta atccccctat ctctcatgga atggactgac
3541 tttatgccta tgaagtcccc aggagctaca ctgatactga gaaaaccagg ctctttgggg
20 3601 ctagacagac tggcagagag tgagatctcc ctctctgaga ggagcagcag atgtctcacg
3661 accacactca gctcaggccc cttggagcag gatggctcct ctaagaatct cacaggacct
3721 cttagtctct gccctatacg ccgccttcac tccacagcct caccctctcc accccatac
3781 tggtagctgt gtaatgagcc aagtggcagc taaaagtgg ggggtgttctg cccagtcccg
3841 tcattctggg ctagaaggca ggggaccttg gcatgtggct ggccacacca agcaggaagc
25 3901 acaaactccc ccaagctgac tcctcctaac taacagtcac gccgtgggat gtctctgtcc
3961 acattaaact aacagcatta atgca

```

SEQ ID NO: 57

Amino acid sequence of human CSFR1 encoded by the DNA sequence shown in SEQ ID NO: 56.

```

30 MGPGLVLLLVATAWHGQGIPVIEPSVPELVVKPGATVTLRCVNGSVBWDGPPSPHWT
YSDGSSSILSTNNATFQNTGTyrCTEPGDPLGGSAAIHLYVKDPAWPWNVLAQEVVVFED
QDALLPCLLTDVPLEAGVSLVRVRGRPLMRHTNYSFSPWHGFTIHRAKFIQSQDYQCSAL
MGGKRVMSISIRLKVKQVIPGPPALTLVPAELVRIRGEAAQIVCSASSVDVNFDFLQHN
NTKLAIPQQSDFHNNRYQKVLTLNLDQVDFQHAGNYSVASNVQGHSTSMFFRVVESAY
35 LNLSEQNLIQEVTVGEGNLKVMVEAYPGLQGFWNTYLGPFSDHQPEPKLANATTKD
RHTFTLSLPRLLKPSEAGRYSFLARNPGGWRALTFFELTLRYPPEVSVIWTFFINGSGTLLCA
ASGYPPQPNVTWLQCSGHTDRCDQAQVLQVWDDPYPEVLSQEPFHKVTVQSLTIVETLEHN
QTYECRAHNSVSGSWAFIPISAGANTHPDEFLLFTPVVACMSIMALLLLLLLLLLYKY
40 KQKPKYQVRWKIIESYEGNSYTFIDPTQLPYNEKWEFPNNLQFGKTLGAGAFGKVV
EATAFGLGKEDAVLKVAVKMLKSTAHADKEALMSELKIMSHLGQHENIVNLLGACTHGGPVL
VITEYCCYGDLLNFLRRKAEAMLGPSLSPGQDPEGVDYKNIHLEBKRYVRRDSGFSSQGV
DTYVEMRPVSTSSNDSFSEQDLKEDGRPLELRDLLHFSSQVAQGMFLASKNCIHRDVA
ARNVLLTNHGVAKIGDFGLARDIMNDSNYIVKGNARLPVKWMAPEIFDCVYTVQSDVWS
YGILLWEIFSLGLNPYPGILVNSKFYKLVKQGYQMAQPAFAPKNIYSIMQACWALEPTHR
45 PTFQQICSFLEQAEQEDRRERDYNLPSSSRSGSGSSSSSELEESSSEHLTCCEQGDIA
QPLIQPNNYQFC

```

SEQ ID NO: 58

gi|6681044|ref|NM\_007779.1| Mus musculus colony stimulating factor 1 receptor (Csfr1), mRNA

```

50 1 cagaactagc agctgggagc cccgtgcccc gccgactctc caacctgcat cggtcacgc
61 taccctctgg aggtatgga gttggggcct cctctggtcc tgcctgtggc cacagtttgg
121 catggtcagg gggccccctg catcgagcct agtggcccag aactggttgt agagccgggt

```

181 gaaacggtga ccctgcatg tgtgagcaat ggcagtgtgg aatgggatgg ccccatctct  
241 cccatctgga ccttgaccc tgaatctccc ggaagcacc tgaccacaag caacgcgacc  
301 ttcaaaaaca ctgggaccta ccgttgtacc gagcttgaag accccatggc aggcagtacc  
5 361 accatccact tgtatgtcaa agatccggcc cactcttggga atttgctggc acaggagggtg  
421 acagtgggtg agggccagga agctgtgctg ccctgtctga tcaactgaccc tgcactgaag  
481 gacagtgtct cactgatgag tgaggggggc aggcaggtct tacgcaaac ggtctacttc  
541 ttctcgccat ggcgaggttc gattatccgc aaggctaaag tccttgacag caatacctac  
601 gtgtgcaaga ccatggtgaa tggtagggaa tccacctcca ctggcatctg gcttaagggtg  
661 aatcgagtc acccagagcc cccacagata aaattggagc cttagcaagct ggtgaggatt  
10 721 cgaggggagg ctgagcagat cgtgtgctcg gccactaac ccgaagtggg attcaacgtt  
781 atcctcaaac gtggagacac caagctggaa atccccctaa acagtgaact ccaagataac  
841 tattataaaa aagtccgggc tctcagcttc aacgctgtgg acttccaaga cgctggcata  
901 tattcttgtg tggccagcaa tgatgttggc acacgcacgg ccaccatgaa ctccagggtg  
961 gtggagagtg cctacttaaa cttgacctct gagcagagcc tcttgacagga ggtgtctgtg  
15 1021 ggtgacagcc tcatcctcac ggtccatgca gatgectacc cttagcataca gcattacaac  
1081 tggacctacc taggtccatt ctttgaagac cagcgcaagc ttgagtttat caccacaagg  
1141 gccatataca ggtacacatt caagctcttt ctgaaccgtg taaaggcctc agaggcgggc  
1201 cagtacttct taatggcaca aaacaaggca ggctggaata atctgacctt tgagctcacc  
1261 ctgcatatc cccagaggt cagtgttaca tggatgcctg tgaatggctc tgatgtcctg  
20 1321 tctgtgagc tctctgggtg ccctcagccc agcgtgacat ggatggagtg cagggggcac  
1381 accgataggt gtgatgaagc ccaggcttgg cacttttga atgacacca cctgaagtc  
1441 ctgagtcaga agcccttcca caaagtgtat attcagagcc agctgcccac tggaccatta  
1501 aaacacaaca tgacttattt ttgcaaaacc cacaacagtg tgggtaacag ctctcagtac  
1561 ttcagggccg tctccctagg acaaagcaag cagctcccg atgagtcct ctctactccg  
25 1621 gtggtgggtg cctgtatgtc tgtcatgtct ctgctggtgc tactgctgtt gctgctcttg  
1681 tacaagtaca agcagaagcc gaagtaccag gtgctgctga agatcatcga gagatacga  
1741 ggcaatagct acaccttcat tgacctact cagttgcctt acaatgagaa gtgggagttc  
1801 cctcggaaca acctgcagtt tggtaagact ctaggagccg gtgctcttgg gaaggtgggtg  
1861 gaggtacag cctttggtct gggcaaaaga gatgcagtgc tgaagggtggc tgtgaagatg  
30 1921 ctaaagtcca cggctcatgc tgatgagaag gaggcctga tgtcagagct gaagatcatg  
1981 agtcacctgg gacagcacga gaatatagtc aacctcttgg gagcctgtac tcacggagga  
2041 cctgtcctgg tctacactga atactgctgc tatggagacc atctcaactt tctccgaagg  
2101 aaggccgagg ctatgctagg acccagcctg agtccctggtc aggactccga gggagactcc  
2161 agctacaaga acatccacct ggagaagaaa tatgtgcgca gggacagtgg ctctccagct  
35 2221 cctcggttag acacctacgt ggagtagagg cctgtctcga ctctctcaag tgactcttc  
2281 tttaaagcaag atctggacaa agagcacagc cggccctgg agctctggga cctgctccac  
2341 ttctccagcc aagtggctca gggcatggcc ttcttctgtt ctaaaaactg catccaccgg  
2401 gacgtagcag ctgaaacgt gctgttgacc agcggacatg tggccaagat tggggacttt  
2461 ggactggcta gggacatcat gaatgactcc aactatgttg tcaagggcaa tgcgctgctt  
40 2521 gtaaagtgga tggcccccaga gagcatcttt gactgctga tcacagttca gagtgtgtg  
2581 tggctcctacg gcatcctcct ctgggagatc ttctcgttg gtctgaaccc ctacccgggc  
2641 atccatgtga acaacaagtt ctacaaactg gtgaaggatg gataccaaat ggcccagcct  
2701 gtatttgcac cgaagaacat atacagcacc atgcagtcct gctgggacct ggagcctacc  
2761 agaagacca ccttccaaca gatctgcttc ctctccagg agcaggcccg actggagagg  
45 2821 agagaccagg actatgctaa cctgccaagc agcgttggca gcagcggcag tgacagtggg  
2881 ggtggcagca gcgttggcag cagcagtgag ccagaagagg agagctccag tgaacacctg  
2941 gcctgctgtg agccagggga catcgcccag cccctgctgc agcctaacaa ctaccagttc  
3001 tgctgaagtg ggaggagag ccgagtcctg ccgctctcta cgtcccgtt ggctcctcc  
3061 atggcagggc gaacatgggg agaacaatat gacttcgccc tcagcttggc ccgactctga  
50 3121 cacttcagaa catgaggggt ctggggaggc cagaggcccc gtttgttccc agagctggg  
3181 ccatcactgc cagtgggggt ctcacagtgc tagcctctat atttactatg ccaactgggtg  
3241 caccctagct tctctttctc catcctattc ccattttaaa aaaccgctc caaactctcg  
3301 tgtttcaatg gaaagactga tttatgtctc aaaagacaag agtctcaaa gctgtgggta  
3361 agctgaaggc ttgctccct gacagatgct tagactacag gcttcttggg acaggtggcc  
55 3421 ccttcctaag ctacaggag tggccaccac tcttgacct cactctgtct atagtcctgc  
3481 cctctgact agcgtgtact gagcgggac ctaaaaagt ttctaccag tgccctgtca  
3541 ctctagactg aaagggtatg ggcctgatgc aaggctgacc acaccaacaa aaaccgtgtg  
3601 ctctctcca agtctgactc gtcctcaatt aatctgtcaa cattaaacta acagtcatta  
3661 acatc

60 SEQ ID NO: 59

Amino acid sequence of mouse CSFR1 encoded by the DNA sequence shown in SEQ ID NO: 58.

5 MELGPPLVLLLATVWHGQGAPVIEPSGPELVVEPGETVTLRCVSNNGSVVEWDGPISPYWTL  
DPESPGSTLTTRNATFKNTGTYRCTELEDPMAGSTTIHLVVKDPAHSWNLLAQEVTVVEG  
QEAVLPCLITDPALKDSVSLMREGGRQVLKRTVYFFSPWRGPIIRKAKVLDSNTYVCKTM  
10 VNGREBSTSTGIWLKVN RVHPEPPQIKLEPSKLVRI RGEAAQIVCSATNAEVGFNVILKRG  
DTKLEIPLNSDFQDNYYKKVRALSNAVD FQDAGIYSCVASNDVGTRTATMNFQVVE SAY  
LNLTSQSL LQEVSVGDSLILTVHADAYPSIQHYNWTYLGPF FEDQRKLEFITQRAIYRY  
TFKLFNLNRVKASEAGQYFLMAQN KAGWNLT FBLTLRYPPEVSVTWMPVNGSDVLFCDVS  
15 GYPQPSVTWMECRGHTDRCD EAAQALQVWNDTHPEVLSQKPF DKV IISQLPIGTLKHNMT  
YFCKTHNSVGNSSQYFRAVSLGQSKQLPDES LFTPVVVACMSVMSLLVLLLLLLLYKYKQ  
KPKYQVRWKI IERYEGNSYTFIDPTQLPYNEKWEFPRNNLQFGKTLGAGAFGKVVEATAF  
GLGKEDAVLKVAVKMLKSTAHAD EKEALMSBLKIMSHLGQHENIVNLLGACTHGGPVLVI  
15 TEYCCYGDLLNFLRRKAEAMLGPSLSPGQDSEGDSSYKNIHLEKKYVRRDSGFSSQGVD T  
YVEMRPVSTSSSDSFFKQDL DKEASRPLEWDLHLHFSSQVAQGM APLASKNCIHRDVAAR  
NVLLTSGHVAKIGDFGLARDIMNDSNYVVKGNARLPVKWMAPE SIFDCVYTVQSDVWSYG  
ILLWBI FSLGLNPYPGILVNNKFKYKLVKDG YQMAQPVFAPKNIYSIMQSCWDLEPTRRPT  
FQICIFLLQE QARLERRDQDYANLPSSGGSSGSDSGGSSGGSSSEPEESSSEHLACCE  
PGDIAQPLLQPNNYQFC

20 SEQ ID NO: 60

gi|34932021|ref|XM\_225897.2| Rattus norvegicus similar to Macrophage colony stimulating factor I receptor precursor (CSF-1-R) (Fms proto-oncogene) (c-fms) (LOC307403), mRNA

1 atgtgcaagg ctgttgtgaa cgctagggaa tccacctcca ttggcatccg gcttaaggtg  
61 aatcgagccc acccagggcc cccacacatc atattggaac ctactaagct ggtgaggatt  
25 121 cgaggggagg ctgccagatc cgtgtgctcg gccactcact cagaagttga attcaacgtt  
181 atcctcaaac gtggagacac caagttggaa atccccataa acagtgactt ccaagacaac  
241 gcttataaaa aggtcctgac tcttaacctc aatgctgtgg acttccaaga tgctggcata  
301 tattcctgtg tggccaacaa cgcagctggc tcgaacacgg ccaccatgaa cttccaggtg  
361 gtggagagtg cctacttaaa cttgacctct gagcagagcc tcttgcaagg ggtgtctgtg  
421 ggtgagaacc tcgacctcac agtcattgca gatgcctacc ctggcctaca gcgttacaac  
481 tggacctacc tagggccggt ctttgaagac ccacacaatc ttgagtttag aacccaagg  
541 accacataca gctactcatt caaactccac ctgaaccgtg taaagccctt ggaggccggc  
601 cgctactcct taatggcaca aaacaaggca ggctggaata atctgacctt tgagctcacc  
661 ctgcgatacc ccccagaagt cagtggtaca tggatacctg tgaacggctc tgatgtcctg  
721 ctctgtgatg tctctgggta tctcagccc aacgtgacat ggatggagtg caggggccac  
35 781 accgataggt gtgatgaggc ccaggcctcg caggtttggg atgacacaca acctgaagtc  
841 ctgagtcaga agcccttcca cagagtgtac cttcagagcc agctgcccac tgggacccta  
901 aagcacaaca tgacttatgt ttgcagagcc cacaacaatg tgggtaacag ctcccagttc  
961 ttcagggcca tctccctagg acaaagcaag cagctccctg atgagtacac gtccactcca  
40 1021 gtggtggtgg cctgtatata tgtcatgtct ctgctggtac tactgctgct gctgctcttg  
1081 tacaagtaca agcagaagcc gaaatatcag gtgcgctgga agatcattga gagctacgag  
1141 ggcaacaact acaccttcat cgacctacc cagttgccct acaatgagaa gtgggagttt  
1201 ccccgaaca acctgaatt tggtgagact ctcgagctg gtgcctttgg gaaggtggtg  
1261 gaggccacag cctttggtct gggcaagaa gatgcagtgc tgaaggtggc tgtgaagatg  
45 1321 ctcaagtcca cggctcatgc cgatgagaag gaggcctaa tgtcagaact gaagatcatg  
1381 agtcacctgg gacagcatga gaacattgtc aacctcttgg gagcctgtac tcatggaggg  
1441 cctgtcctgg tcatcaccga atactgtgc tacggagacc ttctcaactt ccttcgaagg  
1501 aaggccgagg ctatgctggg acccagcctg agtcctggtc aggaccccga gggggactcc  
1561 agctacaaga acatccacct ggagaagaaa tatgtgcgca gggacagtgg cttctccagt  
50 1621 cagggcgtag atacctatgt ggagatgagg cctgtttcga cttcctcaaa tgactccttc  
1681 ttaagcaag atctggacaa agaggccagc cggccgctgg agctctggga cctgcttcac  
1741 ttctctagcc aagtggctca gggcatggct ttccttgctt ctaaaaatg catccatcgt  
1801 gatgtagctg ctcgaaatgt gctgttgacc agcggacatg tggccaagat tggccgactt  
1861 gggctggcta gggacatcat gaatgactct aactatgttg tcaagggcaa tggccgcctg  
55 1921 cctgtaaagt ggatggcccc agagagcatc ttcgactgcg tctacacagt tcagagtgat

1981 gtgtgggtcct acggcctcct cctctgggag attttctcac ttggtctgaa cccctaccca  
 2041 ggcattcttag tgaacaacaa gttctacaaa ctggtgaagg atggatacca aatggcccag  
 2101 cctgtatttg caccggagaa catatacagc atcatgcagt cctgctggga cctggagcct  
 2161 accaaaagac ctaccttcca gcagatctgc ttectcctcc aggaacaggc ccgactggag  
 5 2221 aggagagagc aggactatgc taacctgcc aagcagcagca gcagcagtag cagcagcagt  
 2281 gacagtgggtg gtggcagtg tggttagcagc agtgagcctg aagaggagag ctccagttag  
 2341 cacctggcct gctgtgagcc aggggacatc gccagcccc tgctgcagcc taacaactac  
 2401 cagttctgct gaagcgggac agcagagtc tgctgccctc cacgtcccag ctgcacctcc  
 2461 tccatggatg ggcgacatgg ggagagcata tgaacttcgt cctcagctcg gccagctct  
 10 2521 gacgcttctg aacatgaggg gttcccagag cctgggccat cactgccagt ggggttctca  
 2581 cagtgtctagc ctctatttac tacaccaact ggtgaaccca tacttcaatt tcttcatcct  
 2641 gttcccactt gaaaaaactg tccccaaact tcgtttcaat ggaaagactg atttgtgtct  
 2701 caaaaagaca ggtctcaggt tgtaggtgag cagaagcttg cctccctgac agaggctcag  
 2761 aactgcaggt tcttggggca ggcggccctt cccaagctca cagactggc gccactctta  
 15 2821 ccttctcttt atctacagtc ccgtctcctc ctggatcttg tacactagga gccagtgagc  
 2881 agctgagagc cagggatgct ttacttagtg ccctgcattc taggctggca ggcagtgagc  
 2941 cttggtgcaa ggctgacaac gccaaagcaa tactgcgtgc tctctccaa actgactcgt  
 3001 cctcattaac agtcaacatt aaactaacag cattaacac

SEQ ID NO: 61

20 Amino acid sequence of rat CSFR1 encoded by the DNA sequence shown in SEQ ID NO: 60.

MELGPPLVLLLATVWHGQGAPVIEPSGPELVVEPGETVTTLRCVSNNGSVEWDGPISPYWTL  
 DPESPGSTLTTRNATFKNTGTYRCTELEDPMAGSTTIHLYVKDPAHSWNLLAQEVTVVEG  
 QEAVLPCLITDPALKDSVSLMREGGRQVLRKTVYFFSAWRGFIIRKAKVLDSENTYVCKTM  
 25 VNGREBTSSTGIWLKVNVRVHPEPPQIKLEPSKLVRIERGEAAQIVCSATNAEVGFNVILKRG  
 DTKLEIFLNSDFQDNYYKKVRALSLNAVDQDAGIYSCVASNDVGTRTATMNFQVVBESAY  
 LNLTSQSLQLQFVSVGDSLILTVHADAYPSIQHYNWTYLGPFEDQRKLEFITORAIYRY  
 TPKLFLNRVKASEAGQYFLMAQNKAGWNNLTPELTLYRPPPEVSVTWMPVNGSDVLFCDVS  
 GYPQPSVTWMECRGHTDRCDQAQALQVWNDTHPEVLSQKPFDKVIIQSQLPITGLKHNMT  
 30 YFCKTHNSVGNSSQYFRAVSLGQSKQLPDESLETPVAVACMSVMSLLVLLLLLLLYKYKQ  
 KPKYQVRWKI IERYEGNSYTFIDPTQLPYNEKWEFFRNQLQFGKTLGAGAFGKVVETAF  
 GLGKEDAVLKVAVKMLKSTAHADKEALMSELKIMSHLGQHENIVNLLGACTHGGPVLVI  
 TEYCCYGDLLNPLRRKAEAMLGPSLSPGQDSEGDSSYKNIHLEKKYVRRDSGFSSQGVDT  
 YVEMRPVSTSSSDSFFKQDLOKEPSRPLELWDLHFFSSQVAQGMFLASKNCIHRDVAAR  
 35 NVLLTSGHVAKIGDFGLARDIMNDSNYVVGKNARLPVKWMAPEIILYCVYTVQSDVWSYG  
 ILLWEIFSLGLNPYPGILVNNKFKLVKDGYSMAQPVFAPKNIYSIMQSCWDLEPTRRPT  
 FQQICFLLQEQARLERRDQDYANLPSSGGSSGSDSGGGSSGGSSSEPEEESSEHLACCE  
 PGDIAQPLLQPNYQFAC

SEQ ID NO: 62

40 gi|23110958|ref|NM\_000396.2| Homo sapiens cathepsin K (pseudosclerosis) (CTSK), mRNA

1 aaattttcca gccgatcact ggagctgact tccgcaatcc cgatggaata aatctagcac  
 61 ccctgatggg gtgcccacac ttgctgccc aaacgaagcc agacaacaga ttccatcag  
 121 caggatgtgg gggctcaagg ttctgctgct acctgtggg agctttgtc tgtacctga  
 181 ggagatactg gacaccact gggagctatg gaagaagacc cacaggaagc aatataacaa  
 45 241 caagggtggat gaaatctctc ggcgtttaat ttgggaaaaa aacctgaagt atatttccat  
 301 ccataacctt gaggtctctc ttggtgtcca tacatatgaa ctggctatga accactggg  
 361 ggacatgacc agtgaagagg tgggtcagaa gatgactgga ctcaaagtac ccctgtctca  
 421 ttcccgagc aatgacaccc ttatatccc agaatgggaa ggtagagccc cagactctgt  
 481 cgactatcga aagaaaggat atgttactcc tgtcaaaaat cagggtcagt gtggttctg  
 50 541 ttgggctttt agctctgtgg gtgcccagg gggccaactc aagaagaaa ctggcaaaact  
 601 cttaaactct agtcccaga acctagtggg ttgtgtgtct gagaatgatg gctgtggagg  
 661 ggggtacatg accaatgcct tccaatatgt gcagaagaac cggggtattg actctgaaga  
 721 tgcttaccga tatgtgggac aggaagagag ttgtatgtac aaccaaacag gcaaggcagc

781 taaatgcaga gggtagagag agatccccga ggggaatgag aaagccctga agagggcagt  
 841 ggcccgagtg ggacctgtct ctgtggccat tgatgcaagc ctgacctcct tccagtttta  
 901 cagcaaaggt gtgtattatg atgaaagctg caatagcgat aatctgaacc atgcggtttt  
 5 961 ggcagtgagg tatggaatcc agaagggaaa caagcactgg ataattaaaa acagctgggg  
 1021 agaaaactgg ggaaacaaag gatatacct catggctcga aataagaaca acgcctgtgg  
 1081 cattgccaac ctggccagct tccccagat gtgactccag ccagccaaat ccacctgtct  
 1141 cttccatttc ttccacgatg gtgcagtgtg acgatgcact ttggaagggg gttggtgtgc  
 1201 tatttttgaa gcagatgtgg tgatactgag attgtctgtt cagtttcccc atttgtttgt  
 1261 gcttcaaatg atccttccta ctttgcctct ctccacccat gacctttttc actgtggcca  
 10 1321 tcaggacttt ccctgacagc tgtgtactct taggctaaga gatgtgacta cagcctgccc  
 1381 ctgactgtgt tgtcccaggg ctgatgtgt acaggtacag gctggagatt ttcacatagg  
 1441 ttagattctc attcacggga ctagttagct ttaagcacc tagaggacta gggtaatctg  
 1501 acttctcact tcctaagttc ccttctatat cctcaaggta gaaatgtcta tgttttctac  
 1561 tccaattcat aaatctattc ataagtctt ggtacaagtt tacatgataa aaagaaatgt  
 15 1621 gatttgtctt ccttctttg cacttttgaa ataaagtatt tatctcctgt ctacagttta  
 1681 ataaatagca tctagtacac at

## SEQ ID NO: 63

Amino acid sequence of human CTSK encoded by the DNA sequence shown in SEQ ID NO: 62.

20 MWGLKVL LLPVVSFALYP E EILDTHWELWKKTHR KQYNNKVDEISRRLIWEK NLKYISIH  
 NLEASLG VHTYELAMNHLGDMTSEEVVQKMTGLKVPLSHSR SNTLYIPEWEGRAPDSVD  
 YRKKG YVTPVKNQGGCGSCWAFSSVGALEGLKKKTGKLLNLS PQNLVDCVSENDGCGGG  
 YMTN A FQYVQKNRGIDSEDAYPYVGQEBSCMYNPTGKAAKCRGYREIPEGNEKALKRAVA  
 RVGPVSVAIDASLTSFQFYSGVYDESCNSDNLNHA VLAVGYGIQKGNKH WIIKNSWGE  
 25 NWGNKG YILMARNKNNACGIANLASFPKM

## SEQ ID NO: 64

Amino acid sequence of human CTSK, a soluble active secreted form derived from SEQ ID NO: 63.

30 THWELWKKTHR KQYNNKVDEISRRLIWEK NLKYISIH NLEASLG VHTYELAMNHLGDMTS  
 EEVVQKMTGLKVPLSHSR SNTLYIPEWEGRAPDSVDYRKKG YVTPVKNQGGCGSCWAFS  
 SVGALEGLKKKTGKLLNLS PQNLVDCVSENDGCGGGYMTN A FQYVQKNRGIDSEDAYPY  
 VGQEBSCMYNPTGKAAKCRGYREIPEGNEKALKRAVARVGPVSVAIDASLTSFQFYSGV  
 YDESCNSDNLNHA VLAVGYGIQKGNKH WIIKNSWGENWGNKG YILMARNKNNACGIANL  
 ASFPKM

## 35 SEQ ID NO: 65

gi|31982432|ref|NM\_007802.2| Mus musculus cathepsin K (Ctsk), mRNA

1 gagccacgct tcctatccga aaagagccta gcgaacagat tctcaacagc aggatgtggg  
 61 tgttcaagtt tctgctgcta cccatggtga gctttgctct gtctccggag gaaatgctgg  
 121 acaccacgtg ggagctatgg aagaagactc accagaagca gtataacagc aaggtggatg  
 40 181 aaatctctcg gcgtttaatt tgggagaaaa acctgaagca aatctctgc cataacctgg  
 241 aggcctctct tgggtgccat acatatgaac tggccatgaa ccacttggga gacatgacca  
 301 gtgaagaagt ggttcagaag atgacgggac tcagaatacc tccctctcga tctacagta  
 361 atgacactct ctatacccca gagtgggaag gcagggtccc agactccatc gactatcgaa  
 421 agaaaaggata cgttactcca gtcaagaacc agggccagtg tggttcctgt tgggctttca  
 45 481 gctctgccgg ggccctggag ggccaactca agaagaaaac tggtaaaactc ttagctctga  
 541 gtccccagaa tcttgtggac tgtgtgactg agaattatgg ctgtggaggc ggctatatga  
 601 ccactgcctt ccaatacgtg cagcagaacg gaggcattga ctctgaagat gcttaccat  
 661 atgtggggcca ggatgaaagt tgtatgtata acgccacggc aaaggcagct aaatgcagag



721 ggtacagaga gattcctgtg gggaacgaga aagccctgaa gagagcagtg gcgcgggtag  
 781 gacccatctc tgtgtccatc gatgcaagct tggcatcttt ccagttttac agcagagggtg  
 841 tgtactatga tgaaaattgt gaccgtgata atgtgaacca tgcagtgttg gtggtgggct  
 901 atggcaccga gaagggaagc aagcactgga taattaaaaa cagctgggga gagagctggg  
 5 961 gaaacaaagg atatgctctc ttggctcgga ataagaacaa cgcctgcggc attaccaaca  
 1021 tggccagctt ccccaagatg tgattccagc cagccagccc atctctcttc agattccttc  
 1081 cttcatggtg caagatattg gtggctttgg aaggaggtg gcattggcgt cctgagaggg  
 1141 acgcagcgat gctaactaag attgttctat ttctcctctc gttggtgctt ccagtgaaca  
 1201 ctctacttcc cttctctctg cccaggggcc tttcttttgt ggacacaaca gggcattttt  
 10 1261 ctgagagttg tggactctgt gctggttagc attggagtcc tccagcaggc tggaggacta  
 1321 aggtgacctt cccgagcccc tgcctctctg atacaccagt gaacatttca gtcttccact  
 1381 gagatgcaca aatctattcg tgattctttg acaaatttac atgatattaa aaaaagtgtt  
 1441 tttcctctct tgtatttgaa ataaagtatc tcattttacaa ttt

SEQ ID NO: 66

- 15 Amino acid sequence of mouse CTSK encoded by the DNA sequence shown in SEQ ID NO: 65.

MWVFKLLLLPMVVSFALSPEEMLDTQWELWKKTHQKQYNSKVDEISRRLIWEKNLQKISAH  
 NLEASLGVHTYELAMNHLGDMTSEEVVQKMTGLRIPPSRSYSNDTLYTPEWEGRPDSID  
 YRKKGYVTFVKNGQCGSCWAFSSAGALEGQLKKTKLLALSPQNLVDCVTENYCGGG  
 20 YMTTAFQYVQNGGIDSEDAYPYVGQDESCMYNATAKAACRGYREIPVGNEKALKRAVA  
 RVGPISVSIDASLASFQFYSRGVYDENCDRDNVNHAVLVVGYGTQKGSKHWI IKNSWGE  
 SWGNKGYALLARNKNNACGITNMASFPM

SEQ ID NO: 67

gi|13928757|ref|NM\_031560.1| Rattus norvegicus cathepsin K (Ctsk), mRNA

25 1 cttgtctgaa aagagcatag acaacagatt ctcaacagca ggatgtgggt gttcaagttt  
 61 ttgtctgtac ccgtgggtgag ctttgcctta tccccggagg aaacgctgga cagcagtggt  
 121 gagctgtgga agaagaccca cgggaagcag tacaacagca aggtggatga aatctctcgg  
 181 cgtttaattt gggaaaaaaa cctgaagaaa atttctgtcc ataactctga ggctctctct  
 241 ggtgcccata cgtatgagct ggccatgaat cacctgggag acatgaccag cgaagaagtg  
 30 301 gttcagaaga tgactggact cagagtccca ccttcgogtt ccttcagtaa tgacactctc  
 361 tataccccag agtgggaagg cagagtccca gactccatcg actatcgaaa gaaaggctat  
 421 gttactccag tcaaaaacca gggccagtgt ggctcctgtt gggctttcag ctctcggggt  
 481 gccctggagg gccaaactca gaagaaaact ggcaaaactc tagctctgag tcccagaat  
 541 cttgtggact gtgtgtctga gaactatggc tgtggaggcg gctatatgac cactgccttc  
 35 601 caatatgtgc agcagaatgg aggcattgac tctgaagacg cttaccgcta tgtggggcag  
 661 gatgaaagtt gtatgtataa cggcacggca aaggcagcta agtcgagagg gtacagagag  
 721 atccctgtgg ggaacgagaa agccctgaag agagcagtggt ctcgggtagg acccgtctct  
 781 gtgtccatcg atgcaagctt gacatcttcc caattttaca gcagagggtgt gtactatgac  
 841 gaaaactgcg accgtgataa tgtgaaccat gccgtgttgg tgggtgggcta tggcacccag  
 40 901 aagggaataa agtactggat aattaaaaac agctggggag aaagctgggg aaacaaaggc  
 961 tatgttctct tggctcggaa taagaacaat gcctgtggca ttaccaacct ggccagcttc  
 1021 cccaagatgt gattoaagcc agccagccca tctgttctca cattccttcc tcaacagtgc  
 1081 aagcgaacgg tggttttggg gtgacacctc tgcctccctt ctctccaccc aaggcccttt  
 1141 tctttgtgga cacaactggg cattttctga gagttgtggc ctctgtgctg atagacgctg  
 45 1201 gagtccctca gcaggctgga gyactaaggt gaccttccca agccctcg

SEQ ID NO: 68

Amino acid sequence of rat CTSK encoded by the DNA sequence shown in SEQ ID NO: 67.

MWVFKLLLLPVVSFALSPEETLDTQWELWKKTHGKQYNSKVDEISRRLIWEKNLKKISVH

NLEASLGAHTYELAMNHLGDMTSEEVVQIMTGLRVPPSRSPSNDTLYTPEWEGRVPSID  
 YRKKGYYTPVKNGQCGSCWAFSSAGALEGQLKKKTGKLLALSPQNLVDCVSENYGCGGG  
 YMTTAFQYVQNGGIDSEDAYPYVGQDESCMYNATAKAAKCRGYREIPVGNEKALKRAVA  
 RVGPFVSVIDASLTSPQFYSRGVYDENCDRDNVNHAVLVVGYGTQKGNKYWI IKNWSGE  
 5 SWGNKGYYVLLARNKNNACGITNLASFPM

SEQ ID NO: 69

gi|4503174|ref|NM\_003467.1| Homo sapiens chemokine (C-X-C motif) receptor 4 (CXCR4), mRNA

```

10      1 gtttgttggc tgcggcagca ggtagcaaag tgacgccgag ggcctgagtg ctccagtagc
      61 caccgcatct ggagaaccag cgggttaccat ggaggggagc agtatataca cttcagataa
      121 ctacaccgag gaaatgggct caggggacta tgactccatg aaggaaccct gtttcctgta
      181 agaaaaatgct aatttcaata aaatcttctc gccaccatc tactccatca tcttcttaac
      241 tggcattgtg ggcaatggat tggtcacatc ggtcatgggt taccagaaga aactgagaag
      301 catgacggac aagtacaggc tgcacctgtc agtggccgac ctctctttg tcatcacgct
      15 361 tcccttctgg gcagttgatg ccgtggcaaa ctggtacttt gggaacttcc tatgcaaggc
      421 agtccatgtc atctacacag tcaacctcta cagcagtgtc ctcatcctgg ccttcacatg
      481 tctggaccgc tacctggcca tcgtccacgc caccaacagt cagaggccaa ggaagctgtt
      541 ggctgaaaag gtggtctatg ttggcgtctg gatccctgcc ctctgtctga ctattcccga
      601 cttcatcttt gccaacgtca gtgaggcaga tgacagatat atctgtgacc gcttctaccc
      20 661 caatgacttg tgggtgggtg tgttccagtt tcagcacatc atgggtggcc ttatcctgcc
      721 tggattgtc atcctgtcct gctattgcat tatcatctcc aagctgtcac actccaaggg
      781 ccaccagaag cgcaaggccc tcaagaccac agtcacctc atcctggctt tcttcgctg
      841 ttggctgcct tactacattg ggatcagcat cgactcctc atcctcctgg aaatcatcaa
      901 gcaagggtgt gagtttgaga acactgtgca caagtggatt tccatcaccc aggccttagc
      25 961 tttcttccac tgttgtctga accccatcct ctatgctttc cttggagcca aatttaaaac
      1021 ctctgcccag cagcactca cctctgtgag cagagggtcc agcctcaaga tcctctccaa
      1081 aggaagcgga ggtggacatt catctgttcc cactgagtct gactcttcaa gtttccactc
      1141 cagctaaccac agatgtaaaa gacttttttt tatacgataa ataacttttt ttttaagttac
      1201 acatttttca gatataaaag actgaccaat attgtacagt ttttattgct tgttggattt
      30 1261 ttgtcttggt tttctttagt ttttgtgaag ttttaattgac ttatttatat aaattttttt
      1321 tgtttcatat tgatgtgtgt ctaggcaggga cctgtggcca agttcttagt tgctgtatgt
      1381 ctctgtgtag gactgtagaa aagggaaactg aacattccag agcgtgtagt gaatcacgta
      1441 aagctagaaa tgatccccag ctgtttatgc atagataatc tctccattcc cgtggaacgt
      1501 ttttctctgt cttaagacgt gattttgctg tagaagatgg cacttataac caaagcccaa
      35 1561 agtggatatag aaatgctggt ttttcagttt tcaggagtgg gttgatttca gcacctacag
      1621 tgtacagtct tgtattaagt tgtaataaaa agtacatgtt aaacttactt agtgttatg

```

SEQ ID NO: 70

40 Amino acid sequence of human CXCR4 encoded by the DNA sequence shown in SEQ ID NO: 69.

```

      MEGISITYSDNYTEEMGSGDYDSMKEPCFREENANFNKIFLPTIYSIIFLTGIVGNGLVI
      LVMGYQKQLRSMTDKYRLHLSVADLLFVITLPFWAVDAVANWYFGNFLCKAVHVIYTVNL
      YSSVLILAFISLDRYLAI VHATNSQRPRKLLAEKVYVGVWIPALLLTIPDFIFANVSEA
      45 DDRYICDRFYPNDLWVVFQFQHIMVGLILPGIVILSCYCI IISKLSHSGHGHQKRKALKT
      TVILILAFACWLPYYIGISIDSFILLEI IKQCEFFENTVHKWISITEALAFFHCCLNPI
      LYAFLGAKFKTSAQHALTSVSRGSSLKILSKGRGGHSSVSTESSESSSFHSS

```

SEQ ID NO: 71

gi|2632100|emb|Z80112.1|MMLCR12 Mus musculus lcr-1 gene

```

1 atggaaccga tcagtatata cactttctgat aactactctg aagaagtggg gtctggagac
61 tatgactcca acaaggaacc ctgcttccgg gatgaaaacg tccatttcaa taggatcttc
121 ctgccacca tctacttcat catcttcttg actggcatag tggcaatgg attggtgatc
181 ctggtcatgg gttaccagaa gaagctaagg agcatgacgg acaagtaccg gctgcacctg
5 241 tcagtggctg acctcctctt tgtcatcaca ctccccttct gggcagttga tgccatggct
301 gactggtaact ttgggaaatt tttgtgtaag gctgtccata tcatctacac tgtcaacctc
361 tacagcagcg ttctcatcct ggcttctcag agcctggacc ggtacctcgc catttgtccac
421 gccaccaaca gtcaaaggcc aaggaaactg ctggctgaaa aggcagttct tgtgggcgtc
481 tggatcccag cctcctcctt gactatacct gacttcatct ttgccgacgt cagccagggg
10 541 gacatcagtc agggggatga caggtacatc tgtgaccgcc ttaccctcga tagcctgtgg
601 atggtggtgt ttcaattcca gcatataatg gtgggtctca tctgcccgg catcgctatc
661 ctctcctggt actgcatcat catctctaat ctgtcacact ccaagggcca ccagaagcgc
721 aaggccctca agacgacagt catcctcctc cttagctttct ttgctgtctg gctgccatat
781 tatgtgggga tcagcatcga ctcttctcag cttttgggag tcatcaagca aggatgtgac
15 841 ttcgagagca ttgtgcacaa gtggatctcc atcacagagg cctcgcctt ctccactgt
901 tgctgaacc ccctcctcta tgccttctc ggggccaaagt tcaaaagctc tgcccagcat
961 gactcaact ccctgagcag aggtctccag ctcaagatcc ttccaaagg aaagcggggg
1021 ggacactctt ccgtctccac ggagtcagaa tctctcagtt ttcactccag ctaaccctta
1081 tgcaaaagact tatataatat atatatatat attgataaag acttttttat gttacaccat
20 1141 tttccagat ataagagact gaccagtctt gtacagtttt

```

SEQ ID NO: 72

Amino acid sequence of mouse CXCR4 encoded by the DNA sequence shown in SEQ ID NO: 71.

```

25 MEPISIIYTSNDYSEEVGSGDYDSNKEPCFRDENVHFNRIFLPTIYFIIFLTGIVGNGLVI
LVMGYQKKLRSMTDKYRLHLSVADLLFVITLPPWAVDAMADWYFGKFLCKAVHIIYTVNL
YSSVLILAFISLDRLYLAIHVATNSQRPRKLLAEKAVYGVWIPALLTIPDFIFADVSQG
DISQGGDRYICDRLYPDSLWMVVFQFQHMVGLILPGIVILSCYCIISKLHSHSKGHQKR
KALKTTVILILAFFACWLPYYVGISIDSFILLGVIKQGCDFESIVHKWISITEALAFFHC
CLNPILYAFLGAKFKSSAQHALNSMSRGSLLKILSKGKRGHSSSVTESESSSFHSS

```

30 SEQ ID NO: 73

gi|17902280|gb|AF452185.1|AF452185 Rattus norvegicus strain Holtzman chemokine receptor CXCR4 (Cxc4) gene, complete cds

```

1 atggaatat acacttcgga taactactcc gaagaagtag ggtctggaga ctatgactcc
61 aacaaggaac cctgcttccg ggatgaaaac gaaaacttca acaggatctt cctgcccacc
35 121 atctatttta tcatcttctt gactggcata gtgggcaatg ggttggtaat cctggtcatg
181 ggttaccaga agaagctgag gagcatgaca gacaagtacc ggctgcacct gtccgtggct
241 gacctcctct ttgtcatcac actccccttc tgggcagttg acgccatggc tgactggtac
301 tttgggaaat ttttatgtaa ggctgtgcat atcatctaca ccgtcaacct ttacagcagt
361 gttctcatcc tggccttcat cagcctggag cgctaccttg ccattgtcca cgccaccaac
40 421 agccagagcg cgaggaagct gctggctgaa aaggccgtct atgtgggtgt ctggatcccc
481 gccctcctcc tgactatccc tgacatcacc ttcgccgatg tcagccaggg ggacggcagg
541 tacatctgtg accgccttta ccccgacagc ctgtggatgg tgggtgtcca gttccagcac
601 atcatggtgg gtctcatcct gccgggcatc gtcacctctg cctgttactg catcatcatc
661 tccaagctgt cacactccaa gggccaccag aagcgcaagg ccctcaagac tacggtcatc
45 721 cttatcctgg ctttctttgc ctgctggcta ccgtattacg tggggatcag catcgattcc
781 ttcatccttt tggaggtcat caagcaagga tgtgagttcg agagcgtcgt gcacaagtgg
841 atctccatca cggagggcct cgccttcttc cactgttgcc tgaaccccat cctctacgcc
901 ttctctgggg ccaaattcaa gagctccgag cagcatgcac tcaattccat gagcagaggg
961 tccagcctca agatccttcc caaagggaaa cgggggtggac actcttccgt ctccacagag
50 1021 tcagaatcct caagttttca ctccagctaa

```

SEQ ID NO: 74

Amino acid sequence of rat CXCR4 encoded by the DNA sequence shown in SEQ ID NO: 73.

5 MEIYTSNDYSEEVGSGDYDSNKEPCFRDENENFNRIFLPTIYFIIFLTGIVGNGLVILVM  
GYQKKLRSMTDKYRLHLSVADLLFVITLPFWAVDAMADWYFGKFLCKAVHIIYTVNLYSS  
VLILAFISLDRLAIVHATNSQRPRKLLAEKAVYGVWIPALLLTIPDIIIFADVSQGDGR  
YICDRLYPDSLWMVVFQFHIMVGLILPGIVILSCYCIISKLSHSGHQKRKALKTTVI  
LILAFFACWLPYYVGISIDSFILLEVIKQGEFESVVKWISITEALAFFHCCLNPILYA  
FLGAKFKSSAQHALNSMSRGSLLKILSKGKRGGHSSVSTESSESSFHSS

SEQ ID NO: 75

10 gi|7524353|ref|NM\_013974.1| Homo sapiens dimethylarginine dimethylaminohydrolase 2 (DDAH2), mRNA

1 ccgcttagac aatgccccgg agccgccaga ccgtcgcgcc cctgccccat cgtagtatat  
61 gagctcgccct acacaaggac ccccgctaaa agccagagct cccagtcccc gaggcttgaa  
121 gacggggact cccttctcca ccaactctgt cctcgggggg tggggcccca gccgagatca  
15 181 cagcgcgaca ggagtggggg tggccgctgg agacaggtga agaaacaaga aaactaagaa  
241 atccgagcgg ttggaggggg agtctgtgtg gatgggatgg ggacgccggg ggaggggctg  
301 ggccgctgct cccatgccct gatccgggga gtcccagaga gcctggcgctc gggggaagggt  
361 gcgggggctg gccttcccgc tctggatctg gccaaagctc aaaggagca cggggtgctg  
421 ggaggtaaac tgaggcaacg actggggcta cagctgctag aactgccacc tgaggagtca  
20 481 ttgccgctgg gaccgctgct tggcgacacg gccgtgatcc aaggggacac ggccctaac  
541 acgcgccctt ggagccccgc tcgtaggcca gaggtcgatg gagtccgcaa agccctgcaa  
601 gacctggggc tccgaattgt ggaaatagga gacgagaacg cgacgctgga tggcactgac  
661 gttctcttca cggcgccgga gtttttctga ggctcttcca aatggacca tcaccgagga  
721 gctgagatcg tggcgacac gtcccggtac ttcgccgtct ccactgtgcc agtctcgggt  
25 781 ccctccacc tgcgcggtct ctgcggcatg gggggacctc gcactgtgtt ggcaggcagc  
841 agcgacgctg cccaaaaggc tgtccgggca atggcagtg tgacagatca cccatatgcc  
901 tccctgacct tcccagatga cgcagctgct gactgcctct ttcttcgtcc tgggttgctt  
961 ggtgtgcccc ctttctctct gcaccgtgga ggtggggatc tgcccaacag ccaggaggca  
1021 ctgcagaagc tctctgatgt caccctggta cctgtgtcct gtcagaact ggagaaggct  
30 1081 ggcgcggggc tcagctccct ctgcttggtg ctcagcacac gccccacag ctgagggcct  
1141 ggccttgggg tactgctggc caggggtagg atagtatagg aagtagaagg ggaaggaggg  
1201 ttagatagag aatgctgaat aggcagtagt tgggagagag cctcaatatt gggggagggg  
1261 agagtgtagg gaaaaggatc cactgggtga atcctccctc tcagaaccaa taaaatagaa  
1321 ttgacctttt aaaaaaaaaa aaaaaaaaaa a

35 SEQ ID NO: 76

Amino acid sequence of human DDAH2 encoded by the DNA sequence shown in SEQ ID NO: 75.

40 MGTPGEGLGRCSHALIRGVPESLASGEGAGAGLPALDLAKAQREHGVGGKLRQRLGLQL  
LELPPEESLPLGPLGDTAVIQGDTALITRPWSPARRPEVDGVRKALQDLGLRIVEIGDE  
NATLDGTDVLTGREGFFVGLSKWTNHRGAEIVADTFRDFAVSTVPVSGPSHLRGLCGMGG  
PRTVVAGSSDAAQKAVRAMAVLTDHPYASLTLPDAAADCLFLRPGLPGVPPFLLHRRGG  
DLPNSQEALQKLSDVTLVPVSCSELEKAGAGLSSLCLVLSTRPHS

SEQ ID NO: 77

45 gi|7949034|ref|NM\_016765.1| Mus musculus dimethylarginine dimethylaminohydrolase 2 (Ddah2), mRNA

1 cgcccccgcc ccatcgctgt acgtgagctc toctaccctg ggaaccctc aaaagcccg

```

    61 gctatccgtc tctccagctt gaggactcag actcccatct tcattaactc tctcctctgg
    121 ggggtggggc aaggctgaaa tcacaaggct gcaggagtgg aggtggccgc tgaccagtga
    181 agtcaagaga aaaaacaaaa cagcttggct ggaaggggct tcggtgtgga tgtgatgggg
    241 acgccggggg aggggctggg tcgttgttcc catgccctga tccggggtgt ccccgagagc
5   301 ttggcatccg gggaaggtgc tggcgctggt ctccggctc tggacctggc taaagctcaa
    361 agggagcatg gagtactagg aggtaaactg aggcaacgac tagggctgca gctgcttgaa
    421 ctgcctcctg aggagtcact gccgctggga ccactgcttg gtgacacggc tgtgatccaa
    481 ggagacacgg ccctaatac aaggccctgg agcccagcac gtaggcctga ggttgatgga
    541 gtgctgtaaag ccctccagga cttgggactc cgaattgtgg agatgggaga tgagaatgag
10  601 acgctggacg gcaccgacgt cctcttcacc ggccgggagt ttttcgtagg cctctccaa
    661 tggaccaatc atcgaggagc tgagatcggt gcagacacgt tccgggactt cgctgtctca
    721 acggtaccgg tctcaggctc ctccgaccta cgcggcctct gtggcatggg gggacctcgc
    781 accgtgggtg ctggaagcag cgaggctgcc caaaaagcag tcagggcaat ggcagcgctg
    841 actgatcacc cctacgcctc ctagacacct ccagatgatg cagctagtga ctgtctcttt
15  901 ctgctgctg gggtgctgg tggcaccctc ttcctcctgc accgcgagg ctctgcagaa
    961 gctctctgat gtcacctggt tacctgtgtc ctgctcagaa ctggagaagg ctggagctgg
    1021 cctcagctcc ctctgcctgg tgctcagcac acgccccac tgctgagggc ctggatttgg
    1081 ggatcccact ggtaggaat agagctgtat agtgggtaga atcagctaag agaggctggg
    1141 tagtctgggg gagatgcccc aggataggga aggacttagt gtggaaaaga atagaagcca
20  1201 ttgggtgagt cctctctgtc aaaaccaata aaataaaatt gaccttttag ataaaaaaaa
    1261 aaaaaaaaaa

```

SEQ ID NO: 78

Amino acid sequence of mouse DDAH2 encoded by the DNA sequence shown in SEQ ID NO: 77.

```

25  MGTPGEGLGRCSHALIRGVPESLASGEGAGAGLPALDLAKAQRHGVLGKKLRQRLGLQL
    LELPPEESLPLGPLLLGDTAVIQDIALITRPWSPARRPEVDGVRKALQDLGLRIVEMODE
    NATLDGTDVLTGREGFFVGLSKWNTNHRGAEIVADTFRDFAVSTVPVSGSSHLRQLCGMGG
    PRTVVAGSSEAAQKAVRAMAALTDHPYASLTLPDDAASDCLFLRPGLPGATPFLHRRGS
    AEAL

```

30 SEQ ID NO: 79

gi|34852173|ref|XM\_215315.2| Rattus norvegicus similar to Ddah2 protein (LOC294239), mRNA

```

    1 ggcttctgtg tggatgtgat ggggacgccg ggggaggggc tgggtcgctg ttcccatgcc
    61 ctgatccggg gtgtccccga gagcttggca tccggggaag gtgctggcgc tggctctccg
35  121 gctctggatc tggctaaagc tcaaagggag catggagtac taggaggtaa actgaggcaa
    181 cgactagggtc tgcagcttct tgaactgcct cctgaggaat cactgccgct gggaccactg
    241 cttggtgaca cggctgtgat ccaaggagac acggctctaa tcacaaggcc ctggagccca
    301 gcgcgtaggc ctgaggttga tggagtccgc aaagctctcc aggacttggg gctcagaatt
    361 gtggagatgg gggatgagaa cgctacgctg gacggcaccg acgtcctctt caccggccgg
40  421 gagttttctg taggcctctc caagtggacc aatcatcgag gagctgagat cgtggcagac
    481 acgttccggg acttcgctgt ctctacggta ccggtctctg gcgcctcgca tctgcgcggc
    541 ctctgtggca tgggaggacc tcgcacggtg gtggctggaa gcagtgaggc tgcccaaaaa
    601 gccgtcaggg caatggcagc actgactgat caccctacg cctctctgac cctcccagat
    661 gacgcagcga gtgactgtct cttctctgct cctgggttgc ctggtaccac acctttcctt
45  721 ctgcaccgcg gaggtgggga cctgcccac agccaggagg ctctgcaaaa gctctctgac
    781 gtcacccctg tacctgtgtc ctgctcgga ctggagaagg ttgagctgg cctcagctcc
    841 ctctgcctgg tgctcagcac acgccccac tgctgagggc tgggtttggg gctccaatta
    901 gctaggaata gagccgtcta gggagtagaa tcaggtaata gaggctgggt agtctgtggg
    961 gatgccccag gatagggaag gacttagtat gggacaaaga ctaggagcca gtgggtgagt
50  1021 cttctctgtc aaaaccaata aaataaaatt ggccttttag at

```

SEQ ID NO: 80

Amino acid sequence of rat DDAH2 encoded by the DNA sequence shown in SEQ ID NO: 79.

MGTFGEGLGRCSHALIRGVPELASGEGAGAGLPALDLAKAQREHGVGGKLRQRLGLQL  
LELPPEESLPLGPLLGD TAVIQGDTALITRPWSPARRPEVDGVRKALQDLGLRIVEMGDE  
5 NATLDGTDVLTFTGREFFVGLSKWTNHRGAEIVADTFRDPAVSTVPVSGASHLRGLCGMGG  
PRTVVAGSSEAAQKAVRAMAALTDHPYASLTLPDDAASDCLFLRPGLPGTTPFLLHRGGG  
DLPNSQEQALQKLSDVTLVPVSCSELEKVGAGLSSCLVLSTRPHC

SEQ ID NO: 81

gi|8922242|ref|NM\_018004.1| Homo sapiens hypothetical protein FLJ10134 (FLJ10134),  
10 mRNA

1 gaagcacatc tggacagctg tggcgccctc ttgcgggccg acgtcagccg agcacgtccc  
61 ccacgtccctc tctttctcgc cacttattat ttattcggtt tcccaaagaa gcgactaggg  
121 acccaagttt aaaaattcct cccccactc aatgcgagac gtggccagat cccatccaac  
181 acacggttta attttcatgg ggctctggga tcaaaagaac agaaacagca acaacaaaag  
15 241 cccagccgct gtctgatttt aagctggcaa agtgggaaaa ataaagtgtt gagtaaacag  
301 accaagttag atcatgggga atttcagagg tcatgccctc cctggaacct tcttttttat  
361 tattggtctt tgggtggtga caaagagtat tctgaagtat atctgcaaaa agcaaaagcg  
421 aacctgctat cttggttcca aaacattatt ctatcgattg gaaatttttg aggggaattac  
481 aatagttggc atggctttaa ctggcatggc tggggagcag tttattcctg gagggcccca  
20 541 tctgatgtta tatgactata aacaaggcca ctggaatcaa ctctgggctt ggcattcatt  
601 caccatgtat ttcttctttg ggctgttggg tgtggcagat atcttatgtt tcaccatcag  
661 ttcacttcct gtgtccttaa ccaagttaat gttgtcaaat gccttatttg tggaggcctt  
721 tatcttctac aaccacactc atggccggga aatgctggac atctttgtgc accagctgct  
781 ggttttggtc gtctttctga caggcctcgt tgccttccta gagttccttg ttcggaacaa  
25 841 tgtactctcg gagctattgc ggtcaagtct cattctgctt caggggagct gggtctttca  
901 gattggattt gtctgtatc cccccagtg aggtcctgca tgggatctga tggatcatga  
961 aaatattttg tttctacca tatgcttttg ttggcattat gcagtaacca ttgtcatcgt  
1021 tggaaatgaat tatgctttca ttacctggtt ggtaaatact agacttaaga ggctctgctc  
1081 ctccagaagt ggacttctga aaaatgctga acgagaacaa gaatcagaag aagaaatgtg  
30 1141 actttgatga gcttcagtt tttctagata aaccttttct tttttacatt gtcttggtt  
1201 ttgtttctcg atcttttgtt tggagaacag ctggctaagg atgactctaa gtgtactgtt  
1261 tgcatttcca atttggttaa agtatttgaa tttaaatatt ttctttttag ctttgaaaat  
1321 attttgggtg atacttcat tttgcacatc atgcacatca tggatttcag gggctagagt  
1381 gatttttttc cagattatct aaagttggat gccacacta tgaaagaaat atttgtttta  
35 1441 tttgccttat agatatgtc aaggttactg ggcttgctac tatttgtaac tccttgacca  
1501 tggaaattata cttgtttatc ttgttgctgc aatgagaaat aaatgaatgt atgtattttg  
1561 gtgc

SEQ ID NO: 82

Amino acid sequence of human DERP7 encoded by the DNA sequence shown in SEQ ID  
40 NO: 81.

MGNFRGHALPGTFFFFIIGLWWTCKSILKYICKKQKRTCYLGSKTLFYRLEILEGITIVGM  
ALTGMAGEQFIPGGPHMLYDYKQGHWNQLLQWHHPTMYFFFGLLGVADILCFTISSLPV  
SLTRLMLSNALFVEAFIFYNHTHGREMLDIFVHQLLVVFLTGLVAFLEFLVRNNVLE  
45 LLRSSLLILQGSWFFQIGFVLYPPSGGPAWDLMDHENILFLTICFCWHYAVTIVIVGMNY  
AFITWJVKSRKRLCSSEVGLLKNAEREQESEEM

SEQ ID NO: 83

gi|31542277|ref|NM\_019631.2| Mus musculus RIKEN cDNA C630002M10 gene  
(C630002M10Rik), mRNA

1 ggccccgcgcg gggaaaccag atgaagcaca tctggacagc tgtgctgaga aagtttgtgg  
 61 gcctcttcca ggctgtccac ccgcaaggca cgtccccccac gtagcctcct gctagccact  
 121 tactacttaa ttttttttcc cccaagaagc aataagcaac ccacgcttga atctttttct  
 181 ctccctcccc cccaccccc atgtgaggcg aggccacatc acattcaaca cagtttagtt  
 5 241 ttcatggggc tttgagatca aaagaacaga aacagcaacc aaagctcagc tgccttctga  
 301 tcctaactga caaagtgggg agagtaaggt gtgcgcaaac aggacaagtt gggatcatggg  
 361 gagtttcaaa ggacatgctc tccttgggag tttcttcttc gccatgggct tttgggtggac  
 421 tatgaagaac atcctgaaat ctgtctacaa aaggcaaaact cgaacctgct acctaaactc  
 481 taaaacatta ttacgtcggg cagagatttg ggaaggagtt gttgtgcttt taatgtctct  
 10 541 cactgggtata gctgggtgaac agtttatctc aggaggacct gccttgatct tgcataaaga  
 601 tggccagtggt aaccagatcc tgggctggca tcacacaacc atgtacttat tctttgggct  
 661 acaggggtata acccaaatca tatgtttcac tactaatgta ctccacttt cctcaagcaa  
 721 gttaatgtta tcaattgcca tctttgtgga gacatttatg ttctacaacc acacacacgg  
 781 tcgggaaatg attgacattt ttgtacacca acttctgggc ttcgttggca cattttcggg  
 15 841 tctgtgtgca ggctgtgagt ttcttggagt tcctctgtaa gaacaaacgca cttctggagc tcttgcggtg  
 901 cagttctctc atgtttcaag gaacctgggt ctggcagatg gcgtttgtgc tgtaccccc  
 961 tagtggaggt gctacatgga acctgtcaga tattcaaaat aaaatgtttc tctcaatgtg  
 1021 cttttgctgg cattatgcat caatccttat cctcatggga gtaaaatag ctttggccaa  
 1081 ctgggttagtc aagtctaggc tgagggaagg ctgcacctca gaagttggac tcctgaagca  
 20 1141 tgctgaccgt gagcaagaat cagaagaaga agtatgatct tgaagtcttt cttgataagc  
 1201 cttctccctt tgcgttgcct ttgttcattg ctttgtttct tgacctctgg tctcaagaac  
 1261 acttctctga ggctgactcc atgtctgttg tacttccagt tttgttaaag tttgtgactt  
 1321 taagtatctt actttcagct ctgaaagaac catgagtgat aaattcactt tttacactgt  
 1381 gcatgccatg taattcaaga ccaatcataa ttgttttcca aagtttagtt tctgttccat  
 25 1441 ttattaaaaa tatttttttt attttccggg tagatacctt caagatttgt ggacttgcca  
 1501 tcaactgtaat acatgacgtg ttgacttgta tttgtctatc ttgttgccac aatggaaaat  
 1561 aaatgaatgc atgcaccttg gtgcagaaac caaaaatcct catttttttt ttcttagtaa  
 1621 agtataccgc accctcacat gatacagaaa aaaatctgca tgtacaaaat tccatttctt  
 1681 tgagactttt ctctatggag agcttgatgc aaaaggtaga gcagaagtct tcttctgcat  
 30 1741 tttcaaaaag taatgatgtt gagacatata gtttcaaaga tgacagagac taggagagag  
 1801 gaattgtggt attcacataa cctagcgttg gtcaactgtc catttgctct gttaatgagg  
 1861 taaaggatta aataacaagc ctttctacaa tcctaggtaa aaagtctttt tgcctgtgg  
 1921 gccaagaaac tcccagatgt gtgtaaaaaa aaaattatgc atattactct ggccctactg  
 1981 ttaaccactc tccatgtttc tcataattag ctcatctttc ttcttgactt tgatcttagt  
 35 2041 taaaaggcca aaagggtggt cttcactcta aattaaagg gttaaaatga cttaataggc  
 2101 atatggaccc tttcttacta tcacatccta tgaatctcaa atggaaacaa gaagagaact  
 2161 aaattaatac aatgttacac atcatgggtg acctgaagaa tgagatgaat ataactctct  
 2221 tgcccaaaaa ctgggagcag tcatcttgct acataagatt ttaaaaagac agatgtgagt  
 2281 acccaaaaat atctcttgag ctgtaagtga gtgtactcca ggatacaagg taccgatatg  
 40 2341 tcatggctct ctgacgatca tgtgtctctg tcatgtctgt cttagattta aggactctat  
 2401 ttttagacca caacattctg tgactgccac atttgagctt caaaggaacc aggaatcagc  
 2461 ctgagctagt tgagacaagt cactgatatt gatgacagat aaggttacac ccggaagttt  
 2521 gaaagccaat gaatccagat tttctgctgt ttttatgaga atacagagat cactacttct  
 2581 ccaggttcaa acccagagaa tacaagtaaa cttcaacca gggagttctc agaactctg  
 45 2641 agtctgagac cagttcagga atgcttccct acagtgtctg aaataaaaaac ctttcttgc

SEQ ID NO: 84

Amino acid sequence of mouse DERP7 encoded by the DNA sequence shown in SEQ ID NO: 83.

50 MGSFKGHALPGSFFFAMGFWWTMKNILKSVYKRQTRTCYLSKTLRLRTEIWEGVVVLLM  
 SLTGIAGEQFISGGPALILHKDQWNQILGWHHTTMYLFFGLQGITQIICFTTNVLPSS  
 SKLMLSIAIFVETFMFYNHTGREMIDIFVHQLLVFVGTFSGLVAFLEFLVKNNALLELL  
 RCSLLMPQGTWFWQMAFVLYPPSGSATWNLSDIQNMFLSMCFCHWHYASILILIGVKYAL  
 ANWLKSLRLKGTSEVGLLKHADREQSEBEEV

55 SEQ ID NO: 85

gi|34868010|ref|XM\_340979.1| Rattus norvegicus similar to Dermal papilla derived protein 7 homolog (19.5) (LOC360708), mRNA

```

      1 gccccgccgg gaaaccagat gaagcacatc tggacagctg tgcgagaaa gtttgtgggc
      61 cctttccagg cctgccacca gcaaggcacg tccccacgt agcctcctgc tagccaetta
5    121 ctacttaaat attttttttt cccaagaagc aatcagcaac ccaagcttga atccttttct
      181 tctcccctat ttgaggcgag gccacatcac atcaacacag ttttaattttc atggggcttt
      241 gcgatcaaaa gaacagaaac ggcaacaaaa gtccagctgc cttctgatcc taactgacaa
      301 agtggggacc cagtaagggtg tgagtaaaaca ggccaagctg ggtcatgggg agtttcatag
      361 gtcatgctct ccctgggact ttcttcatca tgatgggctt ttgggtggact acaaagaaca
10   421 ttttgaatc tgtttacaaa aaacacactc gaacctgcta tttgaattct aaaacattat
      481 tacgtcgaat agagatttgg gaaggagttg ttgtggttat aatggctctt actggtatag
      541 ctgggggaaca gtttatctcg ggaggacctg ccttgatctt gtataaagac ggccaatgga
      601 accagatcct gggctggcat cacaccacca tgtacttctt ctttgggtcta cagggtgtaa
      661 ccagatcgt atgttttact actaatgcac ttccgcttct ctttaagcaag ttgatgttag
15   721 cgaatgccat ctttgtggag acatttatct tctacaacca cacacatggt cgggaaatgg
      781 ttgatatttt tgtacaccaa cttctgtcct acaccaccac ggcgggcggt ctagttgcct
      841 tcatggagtt cctcacaagg aacaatgtac ttctggagct cgtgaggtca agtctcatcc
      901 tattacaagg aacctggttc tggcaggttg cttttgttct gtacctcct aaaggaagag
      961 ctacatggaa cctgtccgat attggaata aaatgtttct ctcaatgtgc ttttgttggc
20  1021 attatgcac aattgtcatg ctcacggag taatatgtgc gttggccaac tgggttagta
      1081 aatctagact taggaaggtc tgcacctcag aagttggact ccttaacat gttgacctg
      1141 aacaagaatc agaagaagaa gtatga

```

SEQ ID NO: 86

Amino acid sequence of rat DERP7 encoded by the DNA sequence shown in SEQ ID NO: 85.

```

25  MGSFIGHALPGTFFIMMGFWWTTKNILKSVYKIKHTRTCYLNSTLLRRIEIWEGVVVVIM
    ALTGIAGEQFISGGPALILYKDGQWNQILGWHTTMYFFFGLGQVTVQIVCFTTNALPLSL
    SKLMLANAI FVETFI FYNHHTHG REMVDIFVHQLLSYTTTAAGLVAPMEFLT KNNVLELV
    RSSLILLQGTWFWQVAFVLYPPKGRATWNLSDIGNKMFLSMCFCHYASIVMLIGVIFAL
    ANWLKSLRLKVKCTSEVGLLKHVDRREQESEEEV

```

30 SEQ ID NO: 87

gi|13376090|ref|NM\_024756.1| Homo sapiens elastin microfibril interfacer 3 (EMILIN3), mRNA

```

      1 aagacaacgt cactagcagt ttctggagct acttgccaag gctgagtgtg agctgagcct
      61 gccccaccac caagatgata ctgagcttgc tgttcagcct tgggggcccc ctgggctggg
35   121 ggctgctggg ggcattgggc caggcttcca gtactagcct ctctgatctg cagagctcca
      181 ggacacctgg ggtctggaag gcagaggctg aggacaccag caaggacccc gttggacgta
      241 actggtgccc ctacccaatg tccaagctgg tcaccttact agctctttgc aaaacagaga
      301 aattcctcat ccactcgag cagccgtgtc cgcagggagc tccagactgc cagaaagtca
      361 aagtcattga ccgcatggcc cacaagccag tgtaccaggt caagcagaag gtgctgacct
40   421 ctttggcctg gaggtgctgc cctggctaca cgggccccaa ctgagagcac cagattcca
      481 tggcaatccc tgagcctgca gatcctggtg acagccacca ggaacctcag gatggaccag
      541 tcagcttcaa acctggccac ctgctgcag tgatcaatga ggtgaggtg caacaggaaac
      601 agcaggaaca tctgctggga gatctccaga atgatgtgca ccgggtggca gacagcctgc
      661 caggcctgtg gaaagccctg cctggttaacc tcacagctgc agtgatggaa gcaaatcaaa
45   721 cagggcacga gttccctgat agatccttgg agcaggtgct gctacccac gtggacacct
      781 tcctacaagt gcattttcag cccatctgga ggagctttaa ccaaagcctg cacagcctta
      841 cccaggccat aagaaacctg tctcttgacg tggaggccaa ccgccaggcc atctccagag
      901 tccaggacag tgcgtggcc agggctgact tccaggagct tggtgccaaa tttgaggcca
      961 aggtccagga gaacactcag agagtgggtc agctgcgaca ggacgtggag caacagctgc
50  1021 acgcccagca ctttaccctg accgctoga tctcagagct ccaagccgat gtggacacca
      1081 aattgaagag gctgcacaag gctcaggagg ccccaggagc caatggcagt ctggtgttgg

```



1141 caacgcctgg ggctggggca aggcctgagc cggacagcct gcaggccagg ctggggccage  
 1201 tgcagaggaa cctctcagag ctgcacatga ccacggcccg cagggaggag gagttgcagt  
 1261 acaccctgga ggacatgagg gccaccctga cccggcacgt ggatgagatc aaggaaactgt  
 1321 actccgaatc ggacgagact ttcgatcaga ttagcaaggt ggagcggcag gtggaggagc  
 5 1381 tgcaggtgaa ccacacggcg ctccgtgagc tgcgcgtgat cctgatggag aagtctctga  
 1441 tcatggagga gaacaaggag gaggtggagc ggcagctcct ggagctcaac ctcacgctgc  
 1501 agcacctgca aggtggccat gccgacctca tcaagtacgt gaaggactgc aattgccaga  
 1561 agctctattt agacctggac gtcattccggg agggccagag ggacgccacg cgtgccctgg  
 1621 aggagaccca ggtgagcctg gacgagcggc ggcagctgga cggctcctcc ctgcaggccc  
 10 1681 tgcagaacgc cgtggacgcc gtgtcgctgg ccgtggacgc gcacaaagcg gagggcgagc  
 1741 gggcgcgggc ggccacgtcg cggctccgga gccaaagtga ggcgctggat gacgaggtgg  
 1801 gcgcgctgaa ggcggccgcg gccgaggccc gccacgaggt gcgccagctg cacagcgcct  
 1861 tcgccgccct gctggaggag gcgctcgccc acgaggcggg gctggccgcg ctcttcgggg  
 1921 aggaggtgct ggaggagatg tctgagcaga cgccgggacc gctgcccttg agctacgagc  
 15 1981 agatccgagt ggccctgcag gacgcctgca gcgggctgca ggagcaggcg ctcggtggg  
 2041 acgagctggt gcgccgagt gacggccctg agcaggcctc ggagcccccg cggccggcag  
 2101 agcacctgga gccacgccac gacgcggggc gcgaggaggc cgccaccacc gccctggccg  
 2161 ggctggcgcg ggagctccag agcctgagca acgacgtcaa gaatgtcggg cgggtgctgcg  
 2221 aggcgagagg cggggccggg gccgcctccc tcaacgcctc ccttgacggc ctccacaacg  
 20 2281 cactcttcgc cactcagcgc agcttgaggc agcaccagcg gctcttccac agcctctttg  
 2341 ggaacttcca agggctcatg gaagccaacg tcagcctgga cctggggaag ctgcagacca  
 2401 tgctgagcag gaaagggaag aagcgcaga aagacctgga agctccccg aagagggaca  
 2461 agaagggaag ggagcctttg gtggacatac gggtcacagg gcctgtgcca ggtgccttgg  
 2521 gcgcggcgct ctgggaggca ggatcccttg tggccttcta tgccagcttt tcagaaggga  
 25 2581 cggctgcctt gcagacagt aagttcaaca ccacatacat caacattggc agcagctact  
 2641 tccctgaaca tggctacttc cgagccccct agcgtggtgt ctacctgttt gcagtgagcg  
 2701 ttgaatttgg cccaggggcca ggcaccgggc agctggtgtt tggaggtcac catcggaactc  
 2761 cagtctgtac cactgggcag gggagtggaa gcacagcaac ggtctttgcc atggctgagc  
 2821 tgcagaaggg tgagcgagta tggtttgagt taaccaggg atcaataaca aagagaagcc  
 30 2881 tgtcggggcag tgcatttggg ggcttccctg tgtttaagac ctgaacccca gccccaatct  
 2941 gatcagacat catggactcg cccagctctc ctccggcctg ggctctggcc aaggatgggc  
 3001 tggaggtcat tcagttggte tgtctcttcc ctggaaacct tctgcaaaga tgggtgtggtg  
 3061 tacgtggctt ccctgtaacc acatggggct tggccatttc tccatgatga gaaggactgg  
 3121 aatgcttctc cgggcaggac atggtcctag gaagcctgaa ccttggtctg gcatgccttc  
 35 3181 tcagacagca cggcctgggc tccaactctt caccacaccc tgtattctac aacttctttg  
 3241 gtgttttgct cctcctgtgg ttggaactct ctgtacaaca ctttaaactt ttctcttgct  
 3301 tctcttctc tctcctctta tcgtatgata gaaagacatt cttccccagg aggaatgttt  
 3361 aaaatggagg caacattttg gccaacattg gaaagcacta gagggcaatg ggattaaacc  
 3421 aacctgcttg gtctctatta gtcagtaatg aagacgacag cctggccaac caagggaag  
 40 3481 gaaattagta tctttagttt cagtcattcc ttgtaggata tggtttagct gtgccccac  
 3541 ctaaaatatc atcttgaatt gtaatcccta taatccccac atcaaggag agatcaggtg  
 3601 gaggtaatg gatcttgggg gcggttcccc catgctgttc ttgtgatag tctcacgaga  
 3661 tctgatgatt ttataagttt gatagtctct cctgtgttca ttctccttc tgccaccttg  
 3721 tgaagatgcc ttggttcctc ttactgtctt gccatgattg taagtttctt gaggcctccc  
 45 3781 cagccatgtg gaacagtgg tcaattaac ctctttcctt tataaatt

SEQ ID NO: 88

Amino acid sequence of human ENDOGLYX1 encoded by the DNA sequence shown in  
 SEQ ID NO: 87.

50 MILSLLFSLGGPLGWGLLGAWAQSSTSLSDLQSSRTPGVWKAEDTSKDPVGRNWCPY  
 PMSKLVTLALCKTEKFLIHSQQPCPQAPDCQKVVMYRMAHKPVYQVKQKVLTSLAWR  
 CCPGYTGNPNCHEHDSMAIPEPADPGDSHQEPQDGPVSFKPGHLAAVINEVEVQBEQEH  
 LGDLQNDVHRVADSLPGLWKALPGNLTAAVMEANQTGHEFPDRSLEQVLLPHVDTFPLQVH  
 FSPWRSFNQSLHSLTQAIRNLSLDVEANRQAI SRVQDSAVARADPQELGAKFEAKVQEN  
 55 TORVGQLRQDVEDRLHAQHFTLHRSISELQADVDTKLRLHKAQEAPGTNGSLVLATPGA  
 GARPEPDSLQARLQQLQRLSELHMTTARREBELQYTLEDMRATLTRHVEIKELYSESD  
 ETFPQISKVERQVEELQVNHTALRELRVILMEKSLIMEENKEEVERQLLELNLTLQHLQG  
 GHADLIKYVKDCNCQKLYLDLDVIREQRDTRALEBETQVSLDERRQLDGSSSLQALQNAV

DAVSLAVDAHKAEGERARAATSRLRSQVQALDDEVGALKAAAAEARHEVROLHSAFAALL  
 EDALRHEAVLAALFGEEVLEEMSEQTGGLPLSYEQIRVALQDAASGLQEALGWDELAA  
 RVTALEQASEPPRPAEHLEPSHDAGREAAATTALAGLARELQSLSDNVKQVGRCCAEAG  
 AGAASLNASLDGLHNLATQRSLEQHQRLFHSFGNFQGLMEANVSLDLGKLQTMLSRK  
 5 GKQKQKDLAPRKRDKEAEPLVDIRVTGPVPGALGAALWEAGSPVAFYASFSEGTAAALQ  
 TVKFNTTYINIGSSYFPEHGYFRAPERGVYLFVSVFEGPGPGTGQLVFGGHHRTPVCTT  
 GQSGSGSTATVPAMAELOKGERVWFELTQGSITKRSLSGTAFFGGFLMPKT

SEQ ID NO: 89

gi|37620146|ref|NM\_153127.2| Mus musculus elastin microfibril interfacer 3 (Emilin3),  
 mRNA

1 ttttctgaca tttcgcttga agaccacatc accagcaaat ttgagagtca cttgtaaggc  
 61 tgagcatcag cacaggagcc cctcaccatg atccccgacac tgcctgctggg ctttgggggtg  
 121 tacctgagct ggggactgct agggctcctgg gcacaggacc ccggtaccaa gttctcccat  
 181 ctcaataggc ccggcatgcc tgaaggctgg agactagggg ctgaggatac cagcagagac  
 15 241 cccatcagac ggaactgggtg tccttaccag aagtccaggc tagtcacctt tgtagctgct  
 301 tgcaaaacag agaaattcct ggtccattca cagcagccat gtccacaggg agccccctgac  
 361 tgccaggagg tccagagtcac gtatcgagtg gcccagaagc cagtgtacca ggtccagcag  
 421 aagggtgctga tctctgtgga ctggcggtgc tgcccagggt tccaggagacc agactgccag  
 481 gaccacaatc ccacagcaaa ccctgagccc acagagccaa gtggtaaaact ccaggagact  
 20 541 tgggactcga tggatggctt tgaacttggt caccctgtcc cagagtttaa tgagattaag  
 601 gtgccacaag aacaacagga aaacctgctt caaaatctcc agaattgatgc ccagtcagta  
 661 gaagatggct ttccaggctc ttgggaagcc ccacccagca acctcacaga tgagatgaca  
 721 gaagccaatc taacagaatt cgagtttctt ggcaggacat cagagcacct gctgcagccc  
 781 catattgatg cattcctgaa agcacacttc agtcccatct ggaagaactt caacgacagc  
 25 841 ttgcacagcc tctcccaggc catcagaaac ttgtctcttg atgtggaggc caatcaccag  
 901 gccatcaaga tgatccagga gggcacagtg gctagggtctg acttccaaga gcttggtgct  
 961 aagtttgagg ccaaggtcca gcagaatagc cagagactgg gccaactgtg gcaggatgtg  
 1021 gaggaccagg tgcatgcccc gcgcccagtc gtgcatcatg cctctctctga tgtccaggct  
 1081 gagggtgagca ccaagttaaa gcagcttgct aagggtcagg aacttccagg ggccaatggc  
 30 1141 ggctctggtga tggcatctgc agcagcgcca gcaaggccag agccagagag cctgcaggcc  
 1201 aggttagggc agctgcagag aaacctctct gctctgcaca tggctactag ccagagggag  
 1261 gaggagttgc agagcaccct caagaacatg gacagcgtcc tgaagcagca cgccgaagag  
 1321 atcaaagagc tctattctga atcggtatg accttcgacc agatcagcaa ggtagagagg  
 1381 cagggtggag agctgctggt gaaccacacc gggcttcgag agctgcgggt gatcctaag  
 35 1441 gaaaagtccc tgatcatgga ggagaacaaa gagagatag agcggaact gttggaactc  
 1501 aaccttacc tgcagcatct gcatgcgggt catgcagacc tcattaaagta tgtcaaggac  
 1561 tgcaactgcc aaagggtcaa ctctgacgtg gatgtcgctc cggaggggcca cagggatgtc  
 1621 atgcacaccc tagaagagac ccaagttagc ctggacgaac agcaccagct agacggtct  
 1681 tctttgcagg ccctgcaaaag cactgtagat gccatgtctt cagcaatgga tgcctataga  
 40 1741 ggagagggtg aacggggccc ggctgaaagg gcacggatac ggagccaact gcgggctctg  
 1801 gatcatgctg tggaaagcgt ccttcacagc cctgctggag gatgccctgc gctgctagct  
 1861 ctgcacgggt ccttcacagc cctgctggag gatgccctgc gctgctagct  
 1921 gactcttcg gggaggagat gatagacag atgtcagagg agggccctcg cctctgcca  
 1981 ctggattatg agcagatccg cctagccctg caggacggcg ccagtgggct acagggaacag  
 45 2041 gcgattggtt gggaggactt gggcaccgg gtggaggcat tggagaaggc cgcagggtggc  
 2101 tttgtggagc agcaccacaca gttggcagag ggacttgagc ccagccacga ctctgggaga  
 2161 gaggaggaag ccatgacttt ggcggagctg gaggcaggaga ttccggcgct gagttctgat  
 2221 gtcaagcaga ttgggcagtg ctgtgagggc tcctgggctg cctccctcaa tagtccctt  
 2281 gaagacctac acagcatgct cttggcagacc cagcagggcc tgagacagca ccggcagctc  
 50 2341 ttccacacacc tcttccagaa ctccaaggt ctgggtggcaa gcaacatcag cctagacttg  
 2401 gggaaagctgc aggccatggt gagtaagaaa gataagaagc aaccgagagg cccaggagaa  
 2461 tcccgaaga gggataagaa gcaagtgggt atgtctacag atgcacacgc caaaggctctg  
 2521 gagctctggg agacaggctc ccctgtggcc ttctatgccg gttcttcaga agggggcact  
 2581 gctctgcaga tgggtgaagtt caacaccaca tccatcaatg tgggcagcag ctactttcct  
 55 2641 gaacatggct acttccgagc tcccaaacgt ggcgtctact tgtttgctgt gaggattaca  
 2701 tttggcccag gcccagggaat ggggcagctg gtatttgaag gtcacacccg ggttccagtc  
 2761 tacagtacgg aacagagggg cgggagcaca gcccaccact tttgctatgg tagagctaca

2821 aaagggtgag agagcgtggt ttgagttaat ccaagggtca gcaaccaaag ggagccaacc  
 2881 aggcactgca tttgggggct tcctgatgtt caagacctga accctgggct cggctcaget  
 2941 tgtatcagac atggtatagc cgcttgctgc tcttagactg aggctctggc cagcaaaggc  
 3001 tggagaacat ctagtggcct agctcttccc tggacacctt ctgcaaagac cctgcgggcca  
 5 3061 gcagcacgta ccttcttctg gacacacagg ttggggaggc cagaatactg ctctctggac  
 3121 tgggcccggg cctggagtgg gagctgggat ttccatgcct tcccaaatgg cagggctcgg  
 3181 cttgtctctg agacacttcc tcagttctac agcttttctt tttctctccc ttgtggttgg  
 3241 aaacttgtac acttacgctt ttttcggtt ccacttttct gtgttagaaa gtcactgttt  
 3301 ctcagaagga atgtctacag tgatgggtt gccatgcaga aagggtcccag attctcttct  
 10 3361 caatgtc

## SEQ ID NO: 90

Amino acid sequence of mouse ENDOGLYX1 encoded by the DNA sequence shown in SEQ ID NO: 89.

MIPTLLLGFGVYLSWGLLGSWAQDPGTFKSHLNRPGMPEGWRLGAEDTSRDPIRRNWCPY  
 15 QKSRLVTFVAACKTEKFLVHSQQPCQAPDCQGVVRVYVAQKPVYQVQKVLISVDWR  
 CCPGFGQPDQCQDHNPTANPEPTPSGKLQETWDSMDGFELGHPVPEFNEIKVPQEQQENL  
 LQNLQNDASVEDGFGSWEAPPSNLTDEMTANLTFEPGRTSEHLLQPHIDAFKHAH  
 FSPWKNFNDLSLHLSQAIRNLSLDVEANHQAIKMIQEGTVARADFQELGAKFEAKVQON  
 20 SQRLLGQLWQDVEDQLHAQRRSVHHLSDVQAEVSTKLKQLVKAQELPGANGGLVMASAAA  
 AARPEPESLQARLQQLQRLNLSALHMTVSQREBELQSTLKNMDSVLKQHAEEIKELYSESD  
 ETFDQISKVERQVEELLVNHTGLRELRLVILMEKSLIMEENKEEIERQLLELNLTLQHLHA  
 GHADLIKYVKDCNCQVRVNSDQVDAPEGHRDVMHTLEETQVSLDEQHQLDGSSLQALQSTV  
 DAMSSAMDAYRGEGERARAERARIRSQRALDHAVEALKTAANGTRKEIRLLHGSPTALL  
 EDALRHQAVLAALFGEEMIDEMSEAPRPLDYEQIRLALQDAASGLQEQAIGWEDLAT  
 25 RVEALEKAAGGFVEQHPQLAEGLEPSHDSGREEEAMTLAELEQEIRRLSSDVKQIGQCC  
 ASWAASLNSLLEDLHSMLLDTQHGLRQHRQLFHNLFQNFQGLVASNISLDLGLQAMLK  
 KDKKQPRGPGESRKRDKKQVVMSTDAHAKGLELWETGSPVAFYAGSSEGATALQMVKFNT  
 TSINVGSYFPEHGYFRAPKRGVYLFVAVSITFGPGPMGQLVFEHHRVPVYSTEQRGGS  
 TAAHFCYGRATKG

## 30 SEQ ID NO: 91

gi|27668203|ref|XM\_224646.1| Rattus norvegicus similar to elastin microfibril interfacer 3; EMILIN-like protein EndoGlyx-1 (LOC306288), mRNA

1 atgatactga cactgctgct gggccttgcg gggtagctga gctggggact gctggggctc  
 61 tgggacacagg accctgggtc cagggttctc aaccctaata ggctcagcgt gcctgaaggc  
 121 tggaggatag gggctgacga cgataccagc agagaccca gccaacggaa ctgggtgtcct  
 181 taccagaagt ccaggctggt cacctttgta gctgcctgca aaacagagaa attcctgggt  
 241 cattcacaac agccatgtcc gcaggcgct cctgactgcc agagagtcaa agtcatgtat  
 301 cgagtggccc agaagccagt gtaccaggtc cagcagaagg tgctgatgtc tgtggactgg  
 361 aggtgctgcc cagggttcca ggggtccagc tgccaggacc atcatccac agcaaaccac  
 421 gagccacag aggcaagtgg taaactccag gagacttggg actcattgga tggctttgaa  
 481 cctgggtacc ctgtcacaga gtttaatgag attaaggcgc cacaggaaca cctgcttcaa  
 541 aatctccaga atgacgcca gccggtagaa gatggctttg caggcccttg ggaggcccca  
 601 tccagcaacc tcacagctgc aatgacagaa gccaatctga cagaatttga gtctcttggc  
 661 aggacatcga aacatctgct gcagcccat atccatgcat tcctggaagc acacttcagt  
 721 cccatctgga agagcttcaa cgagcctta cagagcctt tccaggccat cagaatctg  
 781 tctcttgatg tggaggccaa tcaccggcc atcaagatga tccaggaggg caccgtggct  
 841 agggctgact tccgagagct tgggtccaag tttgaggcca aggtccagca gaatagccag  
 901 agactgggac agctgtggca ggatgtggag gaccagctgc atgcccagcg ccgatccgtg  
 961 catcatgccc tctctgaggt ccaggctgag gtgagcacca agttaagca gcttggttaa  
 1021 gctcaggaaac ttccaggggc caacggcagc ctggcggtgg catctgcggc aaggccagag  
 1081 ccagagagcc tgcaggccag actggggcag ctgcagagaa acctctccgc tctgcacatg  
 1141 gtcactaacc agagggagga ggagtgtcgg ggcaccctca aggacatgga cagcgtcctg

```

1201 aggcagcaca cggatgaaat caaggagctc tattctgaat cggatgagac cttcgaccag
1261 atcagcaaag tagagaggca ggtggaggag ctgctggtga accacacggg gcttcgagag
1321 ctacgagtga tcttaatgga gaagtccctg atcatggagg agaacaaga ggagatggag
1381 cggcaactct tggacctcaa cctcactctg cagcatctgc aggcagctca tgcagatctt
5 1441 atcaaataatg tcaaggactg caactgccga aggggtctact ccgacatgga tgtcatccccg
1501 gaggggccgca gggatgccat gcacacccta gaggagaccc aagtaagccc ggacgaacag
1561 caccagctag acgggtcttt gcaggccttg caaagcactg tagacgccat gtcttcggca
1621 ttgatgcct acagaggaga gggatgaacgg gcccgggctg agagggccccg gatgcggagc
1681 cagctgctgg ccctggacca cgtctgtgaa gactgaaga ccgcgccgaa cgggacccgc
10 1741 aaagagatac gcctgctgca tggctccttc gcagccctac tggaggatgc actgcgacat
1801 caggccgtgc tggccgcact cttcggggag gagatggtgg acgagatgtc ggaggagcct
1861 cctcgccctc tggccctgaa ttatgagcag atccgcctgg cctgcagga cgcgcagcgt
1921 gggctgcaag agcaggcggg tgggtgggag gacttggcca ctcggttggg ggctttggag
1981 aaggccgcag gtggctttgt ggagcagcac ccacggctcg cagagggact tgagcccagc
15 2041 caccagcttg ggagagagga ggacaccagg gctttggggg acctggagca agagattcag
2101 cgctgagct cggatgtcaa gcagatgggg cagtgtctgc aggcctcctg ggcctcaat
2161 ggctcccttg aagacctaca cagcgctctc tctgacaccc agcacagcct gagacagcac
2221 cagcagctct tccgtggcct cttccacaac ttccaagggc tgttggcaag caacaccagc
2281 ctagacctgg ataagctgca ggccatgctg agtaagaaag acaagaagca acagaaaggc
20 2341 ccggggagaat cccggaagag ggataagaaa caagtagtga tgtctgcaga tgcctaaagg
2401 ctggggctct gggaagcagg ttccctctgt gccttctatg ccagttcttc agaagaggcc
2461 accgctctgc agatggtgaa gttcaacgct acatccgtca acgtgggcag cggctacttc
2521 cccgaacacg gctatttccg agctcccaaa cgtggcatct acctgtttgc agtgagcgtt
2581 acatttggtc caggccccgg aatggggcag ctgggtattg aagggcacgc ccgggttcca
25 2641 gtctacagt cggaacagag ggggtgggag acagccacta cctttgctat ggcagagctg
2701 caaaaggggt agaggggtgt gtttgagtta attcaagggt cgggtgacaaa ggggagccgg
2761 ccaggcactg catttgagg cttcctgatg ttcaagacct ga

```

SEQ ID NO: 92

30 Amino acid sequence of rat ENDOGLYX1 encoded by the DNA sequence shown in SEQ ID NO: 91.

```

MILTLLGLAGYLSWGLLGSWAQDPGSRFSNPNRSLVPEGWRIGADDDTSRDPSQRNWCP
YQKSRLVTFVAACKTEKFLVHSQQPCPQGAPDCQVRVKMYRVAQKPVYQVQKVLMSVDW
RCCPGFQGPDCQDHHPTANPEPTEASGKLQETWDSLDGFEPGHPVTEFNEIKAPQEHLLQ
NLQNDAPQVEDGFAGPWEAPSSNLTAAMTEANLTFESLGRSTKHLLQPHIHAFLEAHFS
35 PIWKSFNESLQSLFQAIRNLSLDVEANHRAIKMIEGTVARADFRELGAKEAKVQONSQ
RLGQLWQDVEDQLHAQRRSVHHAALSEVQAEVSTKLKQLVKAQELPGANGSLAVASARPE
PESLQARLQQLQRLNLSALHMTNQRBEELRGTCLKMDSVLRQHTDEIKELYSESETFDQ
ISKVERQVEELLVNHTGLRELRLVILMEKSLIMEENKEEMERQLLDLNLTLQHLQAAHADL
IKYVKDCNCRRVYSDMDVIPEGRRDAMHTLEETQVSPDEQHQLDGLQALQSTVDAMSSA
40 LDAYRGEGERARAERARMRSQRLALDHAVEALKTAANGTRKEIRLLHGSFAALLEDALRH
QAVLAALFGEEMVDEMSEEPRLPLNYEQIRLALQDAASGLQEQAQVWEDLATRVEALE
KAAGGFVEQHPRSAEGLEPSHDFGREEDTRALGDLEQEIQRLLSSDVKQMGQCEASWALN
GSLEDLHSALSDTQHSRLRQHQLFRGLFHNFGLLASNTSLDLKQLQAMLSKKDKKQKQG
PGESRKDKKQVMSADAKGLGLWEAGSPVAFYASSSEATALQMVKFNATSVNVGSGYF
45 PEHGYFRAPKRGIIYLFVSVTFPGPGMGQLVFBGHRVPVYSABQRGGSTATTTFAMAE
QKGERVWFELIQGSVTKGSRPGTAFGGFLMFKT

```

SEQ ID NO: 93

gi|37540060|ref|XM\_351265.1| Homo sapiens EGF, latrophilin and seven transmembrane domain containing 1 (ELTD1), mRNA

```

50 1 atgcaccaat tttctgcag atgtaaggaa gggctacagg gtgtcactct gggagctgcg
61 cctgggggata aactgattcc tcagtggggc tgtaacattg ctcagggtca gtggctacta
121 ggcaggggaat ggagcccagg aggcgtgggg aaaggggatg gaggccggac gtggccagcg

```

181 gattcagggc ttgactcaca cactcacaca aaagaggccg gaaagtgtcc ggaggaagcc  
 241 ggcgcgtcac ggaggacgga tccgggaccc tgccgcgcc gccagcgcc gcccggtccc  
 301 ggggtccacag ccgcactcac tccgcgcgc tctccgccac cgcaccact gcggccaccg  
 361 ccaatgaaac gcctcccgt cctagtgggt tttccactt tgttgaattg ttcctatact  
 5 421 caaaattgca ccaagacacc ttgtctccca aatgcaaaat gtgaaatag caatggaatt  
 481 gaagcctgct attgcaacat gggattttca ggaaatgggt tcacaatttg tgaagatgat  
 541 aatgaatgtg gaaatttaac tcagtcctgt ggcgaaaatg ctaattgcac taacacagaa  
 601 ggaagtattt attgtatgtg tgtacctggc ttcagatcca gcagtaacca agacagggtt  
 661 atcactaatg atggaaccgt ctgtatagaa aatgtgaatg caaactgcc a ttagataat  
 10 721 gtctgtatag ctgcaaatat taataaaact ttaacaaaaa tcagatccat aaaagaacct  
 781 gtggctttgc tacaagaagt ctatagaat tctgtgacag atctttcacc aacagatata  
 841 attacatata tagaaatatt agctgaatca tcttcattac taggttaca gaacaacact  
 901 atctcagcca aggacaccct ttctaactca actcttactg aatttgtaaa aaccgtgaat  
 961 aattttggtc aaagggtatc atttgtagt tgggacaagt tatctgtgaa taccaggaga  
 15 1021 acacatctta caaaactcat gcacctgtt gaacaagcta ctttaaggat atcccagagc  
 1081 ttccaaaaga ccacagagtt tgatacaaat tcaacggata tagctctcaa agttttcttt  
 1141 ttgtattcat ataactgaa acatattcat cctcatatga atatggatgg agactacata  
 1201 aatatatttc caaagagaaa agctgcatat gattcaaatg gcaatgttgc agttgcat  
 1261 gtatattata agagtattgg tctttgtct tcatcatctg acaactctt attgaaacct  
 20 1321 caaaattatg ataattctga agaggaggaa agagtcatat cttcagtaat ttcagtctca  
 1381 atgagctcaa acccaccac atttatagaa cttgaaaaaa taacatttac attaatcat  
 1441 cgaaagggtc agataggta taggagtcta tgtgcatttt ggaattactc acctgatacc  
 1501 atgaatggca gctgggtctc agagggtctg gagctgacat actcaaatga gaccacacc  
 1561 tcatgccgct gtaatcacct gacacatttt gcaattttga tgcctctgg tcttccatt  
 25 1621 ggtattaaag attataatat tcttacaagg atcactcaac taggaataat tatttcactg  
 1681 atttgccttg ccatatgcat tttaccttc tggttcttca gtgaaattca aagcaccagg  
 1741 acaacaattc acaaaaatct ttgctgtagc ctatttcttg ctgaacttgt tttcttgtt  
 1801 gggactcaata caaatactaa taagctcttc tgttcaatca ttgccggact gctacactac  
 1861 ttcttttttag ctgcttttgc atggatgtgc attgaaggca tacatctcta tctcattgtt  
 30 1921 gtgggtgtca tctacaacaa gggatttttg cacaagaatt tttatatctt tggctatcta  
 1981 agcccgaccg tggtagttgg attttcggca gcactaggat acagatatta tggcacaacc  
 2041 aaagtatgtt ggcttagcac cgaaaacaac tttatttgga gttttatagg accagcatgc  
 2101 ctaatcattc ttgttaatct cttggctttt ggagtcacat tatacaaaagt tttcgtcac  
 2161 actgcagggt tgaaccagga agttagttgc tttgagaaca taaggtcttg tgcaagagga  
 35 2221 gccctcgctc tctgttctc tctcggcac acctggatct ttgggttct ccattgtgtg  
 2281 cagcatcag tggttacagc ttacctctc acagtcagca atgctttcca ggggatgttc  
 2341 atttttttat tctgtgtgt tttatctaga aagattcaag aagaatatta cagattgttc  
 2401 aaaaatgtcc cctgtgtgt tggatgttta agctgttgaa atgaagtctg ccaaatcttg  
 2461 ctctaacaaa taaaatgtta tctaaatgaa

40 SEQ ID NO: 94

Amino acid sequence of human ETL encoded by the DNA sequence shown in SEQ ID NO: 93.

MHQFFCRCKEGLQGVTLGAAPGDKLIPQWGCNIAQQQWSLGREWSPGGVVGKDGGRTPWA  
 DSGLDSTHTTKEAGKPEEAGASRRTPGPRARPAAPGPGSTAALTTPRSPPPPPLRPP  
 45 PMKRLPPLLVPFSTLLNCSYTQNTKTPCLPNAKCEIRNGIEACYCNMGFSGNGVTICEDD  
 NECGNLTQSCGENANCTNTEGSYYCMCVPGFRSSSNQDRFITNDGTVCIEVNNANCHLDN  
 VCIAANINKTLTKIRSYKEPVALLQEVYRNSVTDLSPTDIITYIEILAESSSLGYKNNT  
 ISAKDTLSNSTLTFVKTVMNFVQRTFVVDKLSVNHRRTHLTKLMHTVEQATLRISQS  
 50 FQKTTEFDNSTDIALKVFFDSYNMKHIHPHMMDGDYINIFPKRKAAYDSNGNVAVAF  
 VYKSGIGPLSSSDNFFLKPNQYDNSEEEERVISVISMSSNPPTLYLEBKITFTLSH  
 RKVTDYRSLCAFWNYSPTMNGSWSEGCETYSNETHTSCRCNHLTHFALMSSGPSI  
 GIKDYNILTRITQLGIIISLILCAICFTFWFFSEIQSTRTTIHKNLCCSLFLAELVFLV  
 GINTNTNKLFCFSIIAGLLHYFFLAFAWMCEGILHYLIVVGVYINKGFLHKNFYIPGYL  
 SPAVVVGFSALGYRYGTTKVCWLSTENNFIWSPFIACLIILVNLLAPGVIIYKVPRH  
 55 TAGLKPEVSCFENIRSCARGALALLFLGTTWIFGVLVVHASVVTAYLFTVSNAPQGMF  
 IFLFLCVLSRKIQEBYYRLPKNVPCCFGLCLSC

SEQ ID NO: 95

gi|18875377|ref|NM\_133222.1| Mus musculus EGF, latrophilin seven transmembrane domain  
containing 1 (Eltl1), mRNA

```

1  acgcgggggac  aggggga'caga  gcgagaccct  gcggtgacctg  cctagctcca  ggacagacccc
5  61  gcactcttagc  ctcgagcgct  ctgcccgcag  ctccgcggct  tccaatgaga  ctctctccgc
121 ttctagtggg  tttctccact  ttgctgaatt  gttcctacac  acaaaactgc  agcaagacaa
181 cgtgtctccc  caatgccaaag  tgcgaagtgc  acaatggtgt  ggaagcctgc  ttctgcagcc
241 aggggtactc  tgggaatggt  gtcacgattt  gtgaagatat  agatgagtgc  agcgagtctt
301 ctgtctgcgg  cgatcatgct  gtgtgtgaaa  acgtgaacgg  gggcttcagc  tgcttctgca
10 361 gggaagggtta  tcagaccgcc  acgggggaag  cacagttcac  acctaatgat  ggctcttact
421 gccaagatat  agatgagtgc  agcgagtctt  ctgtctgcgg  tgatcatgct  gtgtgtgaaa
481 acgtgaacgg  gggcttcagc  tgcttctgca  gggaagggtta  tcagaccgcc  acgggggaag
541 cacagtccac  acctaatgat  ggctcttact  gccaagaaag  catgaattca  aattgccact
601 tagagcatgc  ctgcatcgct  gcaaacatta  ataaaacttt  aaaaagaatt  ggaccataaa
15 661 cagaacagac  aactttactc  caagaaatct  acagaaatc  tgaggctgag  ctctctctga
721 tggatatagt  cacatacata  gagatcctaa  ctgaatcacc  ctactacta  ggccaccoga
781 acagcaccac  ttcatacaag  gatgccact  tcaactcaac  tcttactgaa  tttggggaaa
841 ccatcaataa  ttttgttgaa  aggagtacac  ataaaatgtg  ggaccagtta  ccgacaaatc
901 acagaagact  tcattctaca  aaactgatgc  aactgctga  gctagtacc  ttacagatcg
20 961 ctcagaacat  ccagaagaat  tctcagtttg  atatgaattc  tactgacttg  gctctcaagg
1021 tttttgcttt  tgattcaact  cacatgaagc  atgtcacc  ccacatgaat  gtggatggag
1081 gctatgtgaa  aatatcccca  aggagaaagg  ctgcacatgg  cacaactggc  aatgtagtag
1141 ttgcattcct  ctgctataag  agcattgggc  ccttgcctac  ctcatctgac  aacttcttac
1201 tggacactca  aaatgataat  tctgaaggaa  aggaaaaagt  catttcttca  gtgatttctg
25 1261 cctcaattag  ctcaaattca  cccacattat  atgaacttga  aaaaattaca  ttacactaa
1321 gtcattgtaa  gctctcagat  aagcaccgga  cccagtgtgc  cttttggaac  tactcagttg
1381 atgccatgaa  caatggcacc  tggctaacgg  agggctgtga  gctgacacac  tcaaacgaca
1441 cccacacctc  ctgcgctgt  agtcacctga  cacactttgc  gattttgatg  tctctactt
1501 cttccattgg  gattaaggat  tataatatcc  tgacgaggat  cactcaactc  gggataatca
30 1561 tctccctgat  ctgcctcgcc  atctgcatct  tcacctctg  gttcttcagt  gaaatccaaa
1621 gcaccaggac  cacgattcac  aagaacctgt  gctgcagcct  ctttcttgca  gaactgttt
1681 ttcttattgg  gatcaacata  aatacgaata  agttggtctg  ctctatcatt  gctggcctgc
1741 tccattactt  ctctctagct  gcctttgcct  ggatgtgcat  cgaaggcatt  caccataacc
1801 tcatcgttgt  cggcgctatc  tacaacaagg  ggtttttaca  caagaacttt  tatatctttg
35 1861 gctatctcag  cccagctgta  gttgttggat  tctcagcatc  tttaggatac  agatattatg
1921 gaaccacgaa  agtatgttgg  ctgagcactg  aaaacaactt  catttgagc  tttataggac
1981 cagcgtgtct  aatcattctt  gtgaatctct  tggcttttgg  agttatcata  taaaaagttt
2041 tccgccacac  tgctggactg  aaaccagaag  ttagttgcta  tgagaacata  aggtcttggtg
2101 ccagagggtgc  cctagccctc  ctcttccttc  tgggtaccac  ctggatcttt  ggggttctcc
40 2161 atgtagtga  tgcattctgt  gtgacagcct  acctcttcac  agtcagcaat  gctttccaag
2221 ggatgttcat  tttcttatc  ctatgcgttt  tgtctagaaa  gattcaagag  gaattattata
2281 gattgttcaa  aaatgtcccc  tgctgctttg  gatgtttaag  ataaacaatg  agaagtcagt
2341 ataattacag  ctgcaatgag  atgaaaattc  caagattcag  ataacctgtg  tggcaaaaaa
2401 tgaagcctgt  ttttattgtt  agtaattaat  ttcaaatcca  tttttctgtt  cacagtataa
45 2461 gagatgtagt  taatgtgaga  taaaattatg  gaccagagag  ctacagtgtg  ttttcttaca
2521 tgacatagtt  agagatatgt  caaaaatagt  actgcagata  tttggaaagt  aattggtttc
2581 tctggagtga  tatcactgtg  cccaaggaaa  gatttctttc  taacacgaga  aatatatgaa
2641 tgtcctcaag  gaaaccactg  gctcgatate  tttgtgactc  atgttgccct  tcaaacgagt
50 2701 tccctaccac  cttagtaatg  agttcctttg  caggaaggag  agcataagag  acgtggaggg
2761 gcagagtatg  aagcagtgc  gaaggcttct  ctgacaagga  attgtcattc  caataaactc
2821 agcttctcta  aacttgatga  gaaaatctca  agataaaata  acgagaaagg  aaatatatcc
2881 tagcagtttg  ggaattgggc  tgaagtaaaa  agccccagat  ctaaaatttg  tacatccatg
2941 tcttctctta  ctcttctaaa  accagagaaa  agccttacaa  ctgacattat  cagagatgga
3001 tgctcttaca  ctaacattag  atttgagtgt  aaaatgtttt  cattccacac  agattaagac
55 3061 tctcaaatatg  tagtcagtaa  aacatagatt  tgtcaaagta  taatactgtt  taatgtctta
3121 gtgaaaagaa  tgtgtgcagt  attttgccta  taataattta  ctgttatgaa  aattaccttt
3181 taatattaaa  tcagtatact  tgaatacttt  gctgttttta  atcttcaaaa  tagtgaattt
3241 catgttgcaa  caggcccttt  taattgactg  tacataaaaag  gtcattataa  atttaaagta

```

```

3301 ttgatgaagt gaattataat tcttttctga tcagaaaata cacaattaaa gcattattta
3361 taacaaataa gaagtcactg agtgctgtag ggggttcaca gtgggtctag ttttagactg
3421 tttctactat ctctcaragt ctattggctc aaatgtatgg ctctatctat tctcttgctc
3481 aaatgaagag gcagattttt tttcagaagt gaggcattgt tctgaacctt cstgaacagc
5 3541 ataattcaat ctactggaca ttgcaatttt aattcttggt ctggtgaatg aagcctgtcg
3601 agacctctcc tgaaaaatga acagtcagct ggatgaagca gccctatcgc tgcctgaccg
3661 agttgttctc tcaggagaga ccactcacct gtcaagaagg gctttgcatt tctagagcct
3721 gtgagatggg acactttgac taaatctctg gatttcttct gtgctaagtc tgtggcccat
3781 gactgccatt gtcattctgg gttgggactg tagaaatagg atatcaaac ctagtctgct
10 3841 caatcagtgg atatgaaact attgcacgtg gtagacagag tgaccctcca caaatatctg
3901 tacgctcctc ttcagagcct gtgbcgtggg tggatggcga ggagcaggag ggccctgtgk
3961 ggagggaggg tagctgacct ccactcagtt ggagcactct cactaccctg aggggaagtca
4021 gcattgctgt gatgtatttc agtgggttgc tcctgggttg gaactctcat tcttagagct
15 4081 acaagacaat aaaattctat tatcaaaagg aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

```

SEQ ID NO: 96

Amino acid sequence of mouse ETL encoded by the DNA sequence shown in SEQ ID NO: 95.

```

20 MRLPLLVGFSTLLNCSYQNCSTKCLPNAKCEVHNGVEACFCSQGYSGNGVTICEDID
ECSESSVCGDHAVCENVNGGFSCFCREGYQTATGKSQFTPNDGSYCQDIDECSESSVCGD
HAVCENVNGGFSCFCREGYQTATGKSQFTPNDGSYCQESMNSNCHLEHACIAANINKTLK
RIGPITEQTTLQEIYRNSEAELESLMDIVTYIEILTBSSSLGHFNSTTSYKDAHFNSTL
TEFGETINNFERSTHKMWDQLPTNHRRLHLTKLMHTAELVTLOIAQNIQKNSQFDMNST
25 DLALKVFAFDSTHMKHAHPHMNVDDGGYVKISPRKAAHGTTGNVVVAFLCYKSIGPLLSS
SDNFLDQTQNDNSEGKEKVISSVISASISSNPPTLYELEKITFTLSHVKLSDKHRTQCAF
WNYSVDAMNNGXWSTEGCELTHTSNDTHTSCRCSHLTHFAILMSSTSSIGIKDYNILTRIT
QLGIIISLICLAICIFTWFFPSEIQSTRTTIHKNLCCSLFLAELVFLIGININTNKLVC
IIAGLLHYFFLAFAWMCIEGIIHLYLVVGVYINKGLHKNFYIFGYLSPAVVVGFSA
30 GYRYYGTTKVCWLSTENNFIWFSFGPACLIILVNLLAFGVIYKVFRTAGLKPEVSCYE
NIRSCARGALALLFLGTTWIFGLVHVHASVVTAYLFTVSNAPQGMFIFLFLCVLSRKI
QEYYRFLFKNVPCCFGCLR

```

SEQ ID NO: 97

gi|11560110|ref|NM\_022294.1| Rattus norvegicus ETL protein (Etl), mRNA

```

35 1 ccacaggctg agactagagt ccaggctgtt tgggtgaagg ggcctggcgg ccggacgtgg
61 cctgcagagt ctgggctgtg cacacattca cacaaaagag gccgggaagt gacaggagga
121 agctgtgctg cacaaggagc tgagcgggac cctgccgcgc ctgccagct ccaggacaga
181 ccccaactct tgccctcagc gctctgcgga gccagccagc tccaccgcgc ttccaatgag
241 actcctcctg cttctagtgg gtctctccac tttgctgaat cactcctaca cacaaaactg
301 caagacaccg tgtctcccaa atgccaaagt tgaggtgttg gacgaagtgg cagcctgctt
40 361 ctgcagtaca ggctacactg ggaatggcat cactgattgt gaagatgtag acgagtgcaa
421 cgagacctcc gtctgcggtg atcacgctgt gtgtgaaaac acgaatggag gatttagctg
481 cttctgcgtg gaaggttatc agacctccac cgggaagacg cagttcacgc ctaatgatgg
541 ctcttactgc caagatgtag acgagtcaa cgagacctcc gtctgcggtg atcacgtctg
601 gtgtgaaaac acgaacggag cttctgcgtg gaaggttatc agacctccac
45 661 cgggaagacg cagttcacgc ctaatgatgg ctcttactgc caagaaattg tgaattcaaa
721 ttgccactta gagcatgact gcattgctgc aaacattaat aaaactctaa aaagaattgg
781 acccataaca gaacagctga ctttactcca tgaaatctac aagaattctg aggctgagct
841 ttctctggtg gatatagtca catacataga gatactaaca gaatcatcct cactacaagg
901 ctacataaag aacaccactt cgcccaagga tgcctacttc ggttcagctc ttactgaatt
50 961 tggaaaaacc gtcaataatt ttgttgaaa gaacacacac gaaatgtggg accagttacc
1021 tacaaatcgt agaagactcc atgtcacaaa actgatgcac gctgctgagc acgtcacctt
1081 acagatctct cagaacatcc agaagaatac tcagtttgac atgaattcta ccgacttggc
1141 tctcaagggt ttcgtttttg attcagttca catgaagcat actcatcccc atatgaatgt

```

```

1201 ggacggaggc tatgtaaaaa tatccccgag gagaaaatct gcatatgacc caaatggcaa
1261 cgtcattgtt gcattcctgt gctataggag cattggcccc ttgctttcct catctgacga
1321 cttcttactg ggcgtcaga gtgacaattc caaaggaaag gagaaggcca tttcttcagt
1381 gatttctgcc tcaattagct caaaccacc cactctgtat gaacttgaaa aaattacatt
5 1441 tacactgagt catgtaaagc tctcagataa gcaccagaca cagtgcgcct tttggaacta
1501 ctcagtcgat gacatgaaca atggcagctg gtcactctgag ggctgtgagc tgacatactc
1561 caacgacacc catacttcct gccgatgtag tcatctgaca cactttgcga ttttgatgtc
1621 cccagtgacc tccattgaag ttaaagatta caatatcctg acgaggatca ctcagctggg
1681 aataatcatc tccctgatct gccctgccat atgcattttc accttctggt tcttcagtga
10 1741 gattcaaacg accaggacca caatccacaa gaatctctgc tgcagcctct tcttgcaca
1801 actagttttt cttgtcggca tcaacataaa cacaacaaag ctgggtctgct ctatcatcgc
1861 tggcctgtct cattacttct tcttagctgc ctttgctggt atgtgcattg aaggcatcta
1921 cctatacctc atcgttggtg ggctcatcta taacaagggg tttttacaca agaacttcta
1981 tatcttggc tatcttagcc cggctgtagt tgttggattc tcggcctctt tgggatacag
15 2041 atattatggt accaccaag tatgttggtt gagcactgaa aacaacttta tctggagctt
2101 catagggccca gcgtgtctaa tcattcttgt taatctcttg gcttttgagg ttatcatata
2161 caaagtgttc cgccacactg ctggactgaa gccagaagtt agttgctacg agaacataag
2221 gtcttcgccc agaggagccc tggccctcct cttccttctg ggtaccacct ggacctttgg
2281 ggttctccac gtagtgcatg catctgttgt gacagcctac ctcttcacag tcagcaacgc
20 2341 tttccaaggg atgtttattt tcttattcct atgtgtttta tctagaaaga ttcaagaaga
2401 atattacaga ttgttcaaaa atgtccctg ctgttttgaa tgtttaagat aaacaacgag
2461 aagacacaat aattatagct gaaatgaaat ggaaattcca agatttcgga tagcctgtgt
2521 gacaaaaatg agcctgcctt cattgttagt aattaatttc aaattcgctt ttctgttcgc
2581 agtataaaag atgtagttaa tgtgagataa aattatgggc cagagagctc ctgtgtgttt
25 2641 tcctacatga catagttaga tatgtcaaaa atagtactgc agatatttgg aaagtaattg
2701 gtttctctgg agtgatatca ctgtgcccaa ggaaagattt ctttctaaca caagaaatag
2761 atgaatgtcc tcaaggaagc gactggcttg atatctttgt gactcatgtt gcctttcaaa
2821 cgagtcctc accaccatag taatgagttc ctttgacaga aggagagtat aagaaacttg
2881 gaggggcaga atatgaagca atggagaagc cttctctgac aagggaattgt cattccaata
30 2941 aaattggctt tctccaaaat tgaagaggaa aaaattttca ggctaaaaata acgaaaaagg
3001 aaatgcatcc tagcactttg ggaattggtc tgaacttaa aggccagac ctaaaattac
3061 tacatccatg ttcttctcta ctgttctaaa ccaaagaaaa accttaaaat ttacagatac
3121 atggatgagt gttctcacat aacatcatat ttgaatgtaa atttttttca ttcttcacag
3181 attaaagact cagcaacata tttggtaaaa cataaatttg tcaactata agactgttca
35 3241 tatctttagt gaaaaaatag aatgtgaagt atttgtcta taatatttta ctgttatgaa
3301 aataatcttt tcatattaga gcagtatact tgaatacttt actgttttta atcttcaaaa
3361 tagtgtgatt catgttgcaa ccagcccttt taattgactg tatttaaaag ggcattataa
3421 atttaacta ttgatgaagt aaattataat ggttttctga tcagaaaata catacttaaa
3481 gcattattta taacaaataa aaagtcactg agcactgcag gggtttcaca gtggatctga
40 3541 tatttttaga ccgtttccta tcacctatca gtctatttac ttaaatgtac agctctacca
3601 attctcttac tcaaaggaag aggcagtatt tttctcagaa gtgagtcatt gttctgtacc
3661 ttctgggaga catgattcga tccattgaac attgtgggtt taattcttgt gctgttgaat
3721 gaagcctgac aagacacctc ctaaaaaatg aaatgtcagc tggatgaagc agccctgcta
3781 ctgcctgact gagttgttct ctcaggaaag accactcacc tgccaagaag cacgttgcac
45 3841 ctctacagat ctcagggttt ctccatgcc aagtcctgtg cccacgagca tcattgtcat
3901 tctaagatgg gactgtagaa ataggatatc aaaacataat ccgttcaatc aatggataag
3961 aaactatcac atgtagttaga cagaataacc cttctcaaat attcatacac tctcttcac
4021 aagctgtggc cgtgggtgat agtgaggagc aggaggtcct gtcaggagga agagttagctg
4081 aggtccactc agttggagaa ggctctcact gtgctggggg aagtcagcat gctgacgatg
50 4141 ttacttttagt ttgggtctct tgttttgac atctcatttc tagagctgta aagacaataa
4201 aattctatta tcaaagccaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
4261 aaaaaaaaaa aaaa

```

SEQ ID NO: 98

Amino acid sequence of rat ETL encoded by the DNA sequence shown in SEQ ID NO: 97.

```

55 MRLLLLLVLGLSTLLNHSYQNCKTPCLPNAKCEVLDEVAACFCSTGYTNGITICBDVDE
CNETSVCGDHAVCENTNGGFSCFCVEGYQTSTGKTQFTPNDSYQCQDVDECNETSVCGDH
AVCENTNGGFSCFCVEGYQTSTGKTQFTPNDSYQCEIVNSCHLEHDCIAANINKTLKR

```



IGPITEQLTLHEIYKNSAEELSLVDIVTYIEILTESSSLQGYIKNTTSPKDAYFGSALT  
 EFGKTVNNFVEKNTHEMWDQLPTNRRRLHLTKLMHAAEHVTLQISQNIQKNTQFDMNSTD  
 LALKVFVFDVHMKHTHPHMNVDDGGYVKISPRRKSAYDPNGNVIVAFLCYRSIGPLLSSS  
 5 DDFFLLGAQSDNSKGKEKVISSVISASISSNPPTLYELEKITFTLSHVKLSDKHQQTQCAFW  
 NYSVDDMNNGSWSSSEGCELTYSNDTHTSCRCSHLTHPAILMSPSTSEVKDYNILTRITQ  
 LGIIISLICLAICIFTFWFFSEIQSTRTTIHKNLCCSLFLAQLVFLVGININTNKLVCISI  
 IAGLLHYFFLAFAWMCIIEGIYLYLIVVGLIYNKGFLHKNFYIFGYLSPAVVVGFSASLG  
 YRYYGTTKVCWLSTENNFIWSFIGPACLIILVNLLAFGVIIYKVRHTAGLKPEVSCYEN  
 10 IRSCARGALLFLGTTTWTFGVLHVHASVVTAYLFTVSNAPQGMFIFLFLCVLSRKIQ  
 EBYRYLRFKNVPCCFECLR

SEQ ID NO: 99

gi|31982926|ref|NM\_023928.2| Homo sapiens acetoacetyl-CoA synthetase (AACS), mRNA

1 cgctgaccca gcccgccagg cgctcctgac cgctcgcttc tccggtecca ggtccccggc  
 61 cctcgccctca gccccggccc ctgggtcccca gccctcgctc cagccccggc cggccgcccgc  
 15 121 cgccatgtcc aaggaggagc gccccggctc ggaggagatc ctggagtgcc aggtgatgtg  
 181 ggagcctgac agtaagaaga acacgcagat ggaccgcttc cgggcccgtc tgggcccgcgc  
 241 ctgcggcctc gcgctggaga gttatgatga cttgtacctt tgggtccgtg agtcatattc  
 301 agacttcttg gcagagttct ggaaattcag tggaaattgc ttctcacgtg tgtatgatga  
 361 ggttggtggc acatcgaaag gaatcgaga tgtccccgag tgggtcaaag gcagtcggct  
 20 421 caactatgca gaaaacctcc tgcggcaca agagaatgac agagttgccc ttacattgc  
 481 aagggaaggc aaagaggaaa ttgtgaaggt gacttttgaa gagctgaggc aagaagtggc  
 541 tttgtttgca gcagcaatga ggaaatggg tgtgaagaaa ggagatcggg ttgttggtta  
 601 ttaccacaac agtgagcacg ctgtcgaggc gatgctggct gcggcaagca ttggtgccat  
 661 ctggagctcc acgtccccgg acttcggtgt gaattggtgt ctggaccggt ttctcaaat  
 25 721 tcagccaaag ctcatcttct ctgtggaggc tgtgtctat aatggcaaag agcacaacca  
 781 catggaaaag ctgcagcagg tgggttaaag cctaccagac ttgaagaaag tgggtggtgat  
 841 tccttatgtg tcctccagag agaacataga cttttcaaag attccaaaca gtgtgttct  
 901 ggatgacttt cttgccaccg gcaccagtga gcaggccccg cagctggagt tcgagcagct  
 961 gcccttcagc caccactgtt tcactatgtt ctcatcgggc accacgggcg caccacagt  
 30 1021 catggtgcat tccgctgggg gcacctcat ccagcatctg aaggagcacc tgctgcacgg  
 1081 caacatgacc agcagtgaca cctcctgtg ctacaccacg gtcggctgga tgatgtggaa  
 1141 ctggatggtg tcccttcttg ccacaggagc ggccatggtc ttgtacgatg gctccccct  
 1201 ggtgcccacg cccaatgtgc tctgggacct ggttgacagg ataggcatca ctgtcctggt  
 1261 aactggggcc aagtggctgt cagtgtgga agagaaggcc atgaagccgg tggaaacca  
 35 1321 cagtctccag atgtccaca cgatcctgtc cactggctcc ccactgaaag cccagagcta  
 1381 cgagtatgtc tacagggtga tcaagagcag catcctcctg ggtccatct caggaggcac  
 1441 cgacatcatc tctgtcttca tgggccacaa tttttctctt cctgtgtata aaggggagat  
 1501 tcaggccccg aacctgggca tggccgtgga agcgtggaac gaggaaggaa aggcggtctg  
 1561 gggagagagc ggcgagctgg tgtgtactaa gccgatccct tgccagccca cacacttctg  
 40 1621 gaacgatgag aacggcaaca agtacaggaa ggcgtatttc tccaaattcc caggtatctg  
 1681 ggctcatggc gactactgca gaatcaacct caagaccggg ggcacgtca tgcttggccg  
 1741 gagtgcaggc accctcaacc ccaacggggt gcggttcggc agctcggaaa tctataacat  
 1801 tgtggaatcc ttcgaggagg tggaggacag cctgtgtgtc cccaggtata acaagtacag  
 1861 ggaggagagg gtgactctct tctgaagat ggcctccggg caagccttcc agcctgactt  
 45 1921 ggttaagagg atccgtgacg ccactccgat gggcttgtct gcgcgacacg tgcccagcct  
 1981 catcctggaa accaagggca tcccgatata gctcaacggc aagaaagtgg aagtggcgt  
 2041 caaacagatc atcgctggaa aagccgtgga gcaaggaggc gttttctcga accccgagac  
 2101 cctggatctg taccgggaca tccctgagct gcagggtctc tgagtcagac tggctggcgt  
 2161 gtcactcagc cgcacccgtg tgcactgtaa cttttgtgtg ctcaagaaat tatacagaaa  
 50 2221 cctacagctg ttgtaaaagg atgtcgcac caagtgttct gtaggcttgg ggagggatcg  
 2281 tttctctgtt ttgttaaate tgggtgggac ctggatcttc cacacgagtg ggattctggc  
 2341 tttctagagc caggagggag tgtctgggcc gcagggtgtg cactgtgtgt agagtgtgtg  
 2401 tctttgcaca cacagtgcag tgggaaacgg ggggctggct ggtgtgaa acagacacac  
 2461 tcttgagcca aggtcttctc ttcaacctcc ccgtccgtt gtcccatctt gctctgtgaa  
 55 2521 ggtgcaaatc ctttcttctc ctcccatct caggctctcc tgttttccct cagggtccag  
 2581 tatgcctttg agcttttagt gttagaaagg aaccocgctg acttgacaca gctttcacag  
 2641 ctggctgcta ggaccggcgg gctgggtgtt cacgtgtgtc tgtgtcatgg atgcaatgca

```

2701 ggccctggag gactgtgcgt caccgcgtca ccagagcgtg cctccggggc agcttccctc
2761 caaggaatga gtggatttca tacaggatct ctttattgca cagactgaat ggctttacat
2821 gtttctaata tgaattagga atgtgaagca gtgggtgtcc acccgtgtcc ctcattgggtg
2881 agccctccag ctgtgagccc aggcagtggt gtcaccaggt gaggaccctc ctcaccagga
5 2941 accgcatccc tgtgctgcct ccacctgaga gttgctaggg ggttcttgct gagatcatgt
3001 catcagcacc cctaagtcaa gtcacgggtt tccatagcca ggcagttggt atgtacaatt
3061 cagttcagcg tatgaacttg tatctctaata ctgatgtcca tttttatatt ttttgaaact
3121 gagcacaatg aaatcctttc ttgaatcatt ttccttttgg attataaaaa tatgggggaa
3181 agtgctatga tgagttttat gcaataaatg tatacatgtg tgcacatgca cccaaaaaaa
10 3241 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
3301 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

```

SEQ ID NO: 100

Amino acid sequence of human FLJ12389 encoded by the DNA sequence shown in SEQ ID NO: 99.

```

15 MSKEERPGRBEILECQVMWEPDSKIKNTQMDRFRAAVGAACGLALESYDDLYHWSVESYSD
FWAEFWKFSGIVFSRVYDEVVDTSKGIADVPEWFKGSRNLNYAENLLRHKENDRVALYIAR
EGKEEIVKVTFEELRQEVALFAAAMRMKGVKKGDRVVGYPNSEHAEVAMLAAS IGAIW
SSTSPDFGVNGVLDLRFSGIQPKLIFSVEAVVYNGKEHNHMEKLQVVKGLPDLKKVVIIP
YVSSRENIDLSKIPNSVFLDDPLATGTSEQAPQLEFEQLPFSHPLFIMFSSGTTGAPKCM
20 VHSAGGTLIQHLKEHLLHGNMTSSDILLCYTTVGWMMWNWVMSLLATGAAMVLYDGSPLV
PTPNVLWDLVDRIGITVLVTGAKWLSVLEKAMKPVETHSLQMLHTILSTGSPLKAQSYE
YVYRCIKSSILLGSISGGTDIISCFMGHNFSLPVYKGEIQARNLGMABWNEBGKAVWG
ESGELVCTKPIPCQPTHFWNDENGNKYRKAYFSKFPGIWAHGDYCRINPKTGGIVMLGRS
DGTLPNGVRFSGSSEIYNIVESFEEVEDSLCVPQYNYREERVILFLKMASGHAFQPDLV
25 KRIRDAIRMGLSARHVPSLILETKGIPYTLNGKKVEVAVKQI IAGKAVEQGGAFSNPETL
DLYRDIPELQGF

```

SEQ ID NO: 101

gi|10436299|dbj|AK024036.1| Homo sapiens cDNA FLJ13974 fis, clone Y79AA1001581, weakly similar to ACETYL-COENZYME A SYNTHETASE (EC 6.2.1.1)

```

30 1 cccttggttcc cgcgcgcgcgc cgtcgtgtgac ccagcccgcc aggcgctcct gaccgtcgtc
61 tcctccggtc ccaggtccccc ggccctcgcc tcagcccgcc cccctgggtc ccagccctcg
121 tcgcagcccc ggcgcgtcgc cgcgcgcctg tccaaggagg agcgcggggc tcgggaggag
181 atcctggagt gccaggtgat gtgggagcct gacagtaaga agaacacgca gatggaccgc
241 ttccggggcg ctgtggggcg cgcctgcggc ctggcgctgg agagttatga tgacttgtac
35 301 cattggtccg ttgagtcata ttcagacttc tgggcagagt tctggaaatt cagtgaatt
361 gtcttctcac gtgtgtatga tgaggttggt gacacatcga aaggaatcgc agatgtcccc
421 gagtggttca aaggcagtcg gctcaactat gcagaaaacc tcctgcggca caaagagaat
481 gacagagtta ccctttacat tgcaaggagg ggcaaaggag aaattgtgga ggtgactttt
541 gaagagctga ggcaagaagt ggctttgttt gcagcagcaa tgaggaaaaat ggtgtgaaag
40 601 aaaggagatc ggggtgttgg ttatttacct aacagtgagc acgctgtcga ggcgatgctg
661 gctgcggcaa gcattggtgc catctggagc tccacgtccc cggacttcgg tgtgaatggt
721 gtgctggacc ggttttctca aattcagcca aagctcatct tctctgtgga ggtgtgtgtc
781 tataatggca aagagcacia ccacatggaa aagctgcagc aggtgggttaa aggcctacca
841 gacttgaaga aagtgggtgt gattccttat gtgtcctcca gagagaacat agacctttca
45 901 aagattccaa acagtgtgtt tctggtagac tttcttgcca cgggcaccag tgagcaggcc
961 cgcagctgg agttcgagca gtgcccctc agccaccac tttctcatcgt gttctcatcg
1021 ggcaccacgg ggcacccaa gtgcatggtg cattccgctg ggggcacct catccagcat
1081 ctgaaggagc acctgctgca cggcaacatg accagcagtg acatcctcct gtgctacacc
1141 acggtcggct ggatgatgtg gaactggatg gtgtcccttc tggccacagg agcggccatg
50 1201 gtcttgtacg atggctcccc octggtgccc acgcccattg tgctctggga cctggttgac
1261 aggataggca tcaactgcct ggttaactgg gccaagtggc tgtcagtgct ggaagagaag
1321 gccatgaagc cgggtgaaac ccacagtctc cagatgtccc acacgatcct gtccactggc

```

```

1381 tccccactga aagcccagag ctacgagtat gtctacaggt gcatcaagag cagcatcctc
1441 ctgggctcca tctcaggagg caccgacatc atctcctgct tcatgggcca caatttttct
1501 cttcctgtgt ataaagggga gattcaggcc cggaacctgg gcatggccgt ggaagcgtgg
1561 aacgaggaag gaaaggcggg ctggggagag agcggcgagc tgggtgtgtac taagccgatc
5 1621 ccttgccagc ccacacactt ctggaacgat gagaacggca acaagtacag gaaggcgtat
1681 ttctccaaat tcccagggtat tgggctcatg gcgactactg cagaatcaac cccaagaccg
1741 ggggcatcgt catgcttggc cggagtgcgc gcaccctcaa cccaacggg gtgcggttcg
1801 gcagctcgga aatctataac attgtggaat ccttcgagga ggtggaggac agcctgtgtg
1861 tccccagta taacaagtac agggaggaga ggggtgacct cttcctgaag atggcctccg
10 1921 ggcacgcctt ccagcctgac ttggttaaga ggatccgtga cgccatccgc atgggcttgt
1981 ctgcgcgaca cgtgccagc ctcacctgg aaaccaaggg catcccgtat acgctcaacg
2041 gcaagaaagt ggaagttgcc gtcaaacaga tcatcgctgg aaaagccgtg gagcaaggag
2101 gtgctttctc gaaccccgag accctggatc tgtaccggga catccctgag ctgcagggtc
2161 tctgagtcag actggctggc gtgtcactca gccgcaccgc tgtgactgt aacttttgtg
15 2221 tgctcaagaa attatacaga aacctacagc tgttgtaaaa ggatgctcgc accaagtgtt
2281 ctgtaggctt ggggagggat cgtttctctg ttttgtaaaa tctgggtggg acctggatct
2341 tccacacgag tgggattctg gccttcagag accaggaggg agtgtctggg ccgcagggtg
2401 ggcactgtgg tgagagtgtg tgtctttgca cacacagtgc agcgggaacg gtggggctgg
2461 ctggtgctga agacagacac actcctgagc caaggtcttg tctcaacct ccccgctccg
20 2521 ttgtcccat ttgctctgtg aagggtcaaa tccctttctt cccttcccat ctgaggctct
2581 cctgttttcc ctgagggtcc agtatgcctt tgagctttag ctgtagaaa ggaaccccg
2641 tgacttgaca cagctttcac agctggctgc taggaccggc gggctgggtg ttcacgtgtg
2701 tctgtgtcat ggatgcaatg caggccctgg aggactgtgc gtcaccgctc aaccagagcg
2761 tgcctccggg ccagcttccc tccaaggaaat ggggtggatt catacaggat ctctttattg
25 2821 cacagactga atggctttac atgtttctaa tgtgaattag gcatgtgaag cagtgggtgt
2881 ccaccctgtg ccctcatggg tgagccctcc agctgtgagc ccaggcagtg tggtcaccga
2941 gtgaggaccc tcctcaccag gaaccgcac cctgtgctgc ctccacctga gaggttgtag
3001 ggggttcttg tcgagatcat gtcactcaga cccctaagtc aagtcacggg ttccatagc
3061 caggcagttg gtatgtacaa ttcagttcag cgtatgaact tgtatctcta atctgatgtc
30 3121 catttttata ttttttgaag ctgagcaca tgaatccct tcttgaatca ttttctttt
3181 ggattataaa aatatggggg aaagtgcctat gatgaatttt atgcaataaa tgtatacatg
3241 tgtgcacatg c

```

SEQ ID NO: 102

35 Amino acid sequence of human FLJ12389 variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 101.

```

MSKEERPGRREEILECQVMWEPDSKKNTQMDRFRAAVGAACGLALESYDDLYHWSVESYS
FWAEFWKPFSEIVFSRVYDEVVDTSKGIADVPEWFKGSRLNYAENLLRHKENDRVALYIAR
EGKEEIVEVTFEELRQEVAFAAAMRKMVGKGDVVGYPNSEHAVEAMLAASIGAIW
40 SSTSPDFGVNGVLDLRFQSIQPKLIFSVEAVVYNGKEHNHMEKLQVVKGLPDLKKVVVIP
YVSSRENIDLSKIPNSVFLDDFLATGTSEQAPOLEFEQLPFSHPFIMPSSGTTGAPKCM
VHSAGGTLIQHLKEHLLHGNMTSSDILLCYTTVGWMMWNWMSLLATGAAMVLYDGSPLV
PTPNVLWDLVDRLGITVLVTGAKWLSVLEEKAMKPVETHSLQMLHTILSTGSPLKAQSYE
YVYRCIKSSILLGSISSGTDIISCFMGHNFSLPVYKGEIQARNLGMAVEAWNBEKAVWG
ESGELVCTKPIPCQPTFWNDENGKRYKAYFSKPGIGLMATTAESTPRPGASSCLAGV
45 TAPSTPTGCGSAARKSITLWNPSRRWRTACVSPSITSTGRRG

```

SEQ ID NO: 103

gi|10434316|dbj|AK022740.1| Homo sapiens cDNA FLJ12678 fis, clone NT2RM4002409, weakly similar to ACETYL-COENZYME A SYNTHETASE (EC 6.2.1.1)

```

50 1 agacacatcg aaaggaatcg cagatgtccc cgagtgggtc aaaggcagtc ggctcaacta
61 tgcagaaaac ctctgcggc acaaagagaa tgacagagtt gccctttaca ttgcaaggga
121 aggcaaagag gaaattgtga aggtgacttt tgaagagctg aggcaagaag tggctttgtt
181 tgcagcagca atgaggaaaa tgggtgtgaa gaaaggagat cgggttgggt gttattttacc

```

```

241 caacagtgcg cacgctgtcg aggcgatgct ggctgcggca agcattgggtg ccattctggag
301 ctccacgtcc ccggacttcg gtgagaatgg tgtgctggac cggttttctc aaattcagcc
361 aaagctcatc ttctctgtgg aggcgtgtgt ctataatggc aaagagcaca accacatgga
5 421 aaagctgcag caggtgggta gaggcctacc agacttgaag aaagtgggtg tgattcctta
481 tgtgtcctcc agagagaaca tagaccttc aaagattcca aacagtgtgt ttctggatga
541 ctttcttgcc accggcacca gtgagcaggc cccgcagctg gagttcgagc agctgccctt
601 cagccacca ctgttcatca tgttctcatc gggcaccacg ggcgcaccca agtgcattgt
661 gcattccgct gggggcacc ccatccagca tctgaaggag cacctgctgc acggaacat
10 721 gaccagcagt gacatcctcc tgtgtacac cacggtcggc tggatgatgt ggaactggat
781 ggtgtccctt ctggccacag gagcgcccat ggtcttgtac gatggctccc cctgggtgcc
841 cagcccaat gtgctctggg acctgggtga caggataggg tatctgggct catggcgact
901 actgcagaat caaccccaag accgggggca tctcatgct tggccggagt gacggcacc
961 tcaaccccaa cggggtgcgg ttccggcagc cggaaatcta taacattgtg gaatccttcg
1021 aggaggtgga ggacagcctg tgtgtccccc agtataacaa gtacagggag gagagggtga
15 1081 tcctcttctt gaagatggcc tccgggcacg ccttccagcc tgacttggtt aagaggatcc
1141 gtgacgccat ccgcatgggc ttgtctgcgc gacacgtgcc cagcctcatc ctggaaacca
1201 agggcatccc gtatacgtc aacggcaaga aagtggaaat tgccgtcaaa cagatcatcg
1261 ctggaaaagc cgtggagcaa ggaggtgctt tctcgaacce cgagaccctg gatctgtacc
1321 gggacatccc tgagctgcag ggcttctgag tcagactggc tggcgtgtca ctacagccga
20 1381 cccgtgtgca ctgtaacttt tgtgtgtcga agaaattata cagaaaccta cagctgttgt
1441 aaaaggtgc tcgcaccaag tgttctgtag gcttggggag ggatcgtttc tctgttttgt
1501 taaatctggg gggtagctgg atcttcaca cgagtgggat tctggccttc agagaccagg
1561 agggagtgtc tgggcccgcg gtgtggcact gtggtgagag tgtgtgtctt tgcacacaca
1621 gtgcagcggg aacggtgggg ctggctggtg ctgaagacag acacactcct gagccaaggt
25 1681 cttgtcttca acctcccggt cccgttgtcc cattttgctc tgtgaagggt caaatccctt
1741 tcttcccttc ccatctcagg ctctcctgtt ttccctcagg gtccagtatg cctttgagct
1801 ttagctgtta gaaaggaatg agtggatttc atacaggatc tctttattgc acagactgaa
1861 tggctttaca tgtttcta atgtgaattagg catgtgaagc agtgggtgtc caccctgtgc
1921 cctcatgggt gagccctcca gctgtgagcc caggcagtgt ggtcaccgag tgaggacct
30 1981 cctcaccagg aaccgcatcc ctgtgctgcc tccacctgag agttgctagg ggttctctgt
2041 cgagatcatg tcatcagcac ccctaagtca agtcacgggt ttccatagcc aggcagttgg
2101 tatgtacaat tcagttcagc gtatgaactt gtatctctaa tctgatgtcc atttttatat
2161 tttttgaaac tgagcacaat gaaatccttt cttgaatcat tttccttttg gattataaaa
2221 atatggggga aagtgtatg atgaatttta tgcaataaat gtatacatgt gtgcacatgc
35

```

SEQ ID NO: 104

Amino acid sequence of human FLJ12389 variant ORF number 2 encoded by the DNA sequence shown in SEQ ID NO: 103.

```

40 MRKMGVKKGDRVVGYLPNSEHAVEAMLAASIGAIWSSTSPDFGENGVLD RFSQIQPKLI
FSVEAVVYNGKEHNHMEKLQQVVRGLPDLKKVVVI PYVSSRENIDLSKIPNSVFLDDFLA
TGTSEQAPQLEFEQLPFSHPLFIMFSSGTTGAPKCMVHSAGGTLIQHLKEHLLHGNMTSS
DILLCYTTVGMMWNWVSLLATGAAMVLYDGSPLVPTPNVLWDLVD RIGYLGSRLLQN
QPQDRGHRHAWPE

```

SEQ ID NO: 105

45 gi|21313519|ref|NM\_030210.1| Mus musculus acetoacetyl-CoA synthetase (Aacs), mRNA

```

1 gagtctcgcg ctgtggttcg tcggcgacac gctgatccgc tccacgcctt gcgctctccg
61 ctctcagcca aagcccgcca gccccggcca cgcagctccg caaccatgtc caagctggcg
121 cggctcagac gcgaggagat catggagtgc caggtgatgt gggagcctga cagcaagaag
181 gacacgcaga tggaccgctt ccggcgccgc gtgggtaccg cctgcggcct ggcgcttggg
50 241 aattacaatg acttatacca ctggtctgtc cggctcgata tggacttttg ggctgagttc
301 tgggaagttca gtggaatcgt ctactcacgc atgtatgatg aggttgtgga cacatccaaa
361 gggattgcag atgtcccccga gtggttcaga ggcagccgcc tcaactacgc agagaacctc
421 ctgcggcaca aggagaacga cagagtcgcc ctttacgtgg cccgggaagg cagagaggag

```

```

481 atcgtgaagg tgacttttga agagctgcgg cagcagggtgg ctctgttcgc agctgccatg
541 aggaagatgg gcgtgaagaa aggggaccgt gtgggtcggct atctacccaa cagtgcgcat
601 gcggtggagg ccatgctggc tgctgccagt attggggcca tttggagttc tacctcgccg
661 gacttttggtg taaatggtgt cctggaccgc ttttctcaaa ttcagccgaa acttatcttc
5 721 tcggtggagg ctggtgtcta caatggcaag gagcatggtc acctggagaa gctgcagcga
781 gttgtgaaag gactgcctga cctgcagcga gtgggtgctga tcccctatgt cctcccaagg
841 gagaagattg acatttccaa gatcccaac agtgtgttcc tggatgattt cctggcaagc
901 gggaccggcg cgcaggcgcc gcagctcgag tttgagcagc tgcccttcag ccacctctg
961 ttcacatcatg tctcctcagg cagcagcgga gcgcccaggt gcatgggtgca ctcagccggg
10 1021 ggcaccctca tccagcacct gaaggagcac atgctacacg gcaacatgac aagcagtgcac
1081 atcctgctct actacaccac ggtcggctgg atgatgtgga actggatggg gtcagccctg
1141 gccacaggag catccttggg cctgtatgat ggctccccgc tgggtccgac ccccaacgtg
1201 ttgtgggacc ttgtggacag gataggatc accatcctgg gaacgggagc caagtggctg
1261 tcagtgtctgg aggagaagga catgaagcca gtggaaactc acaacctcca cagctgcac
15 1321 acgatcctgt ccaccggctc gccgctgaaa gccagagtt acgagtatgt gtacagatgc
1381 atcaagagct ccgtgctcct gggctccatc tcaggaggca ctgacatcat ctctgtttc
1441 atgggccaga actcctctat tctgtgtac aagggtgaga tccaagcccg gaaccttggc
1501 atggctgtgg aagcctggga cgaggaaggg aaggccgtct ggggagcgag tggcgagctg
1561 gtgtgcacta agcccatctc ctgccagccc acgcacttct ggaacgacga gaacggcagc
20 1621 aagtaccgga aggcctactt ctccaaattc ccagggtgtc gggcacacgg tgactactgc
1681 aggatcaacc ccaaacaggg aggcattatc atgctgggct gtagtgatgg caccctcaac
1741 ccaaatggcg tccgctttgg cagctcggag atctacaaca tcgtggaagc cttcagatgag
1801 gtggaggaca gcctgtgtgt accccagtae aacagagatg gcgaggagcg ggtggtcctg
1861 ttctgaaga tggcgctcgg gcacactttc cagcctgacc tcgtgaagcg catccgagac
25 1921 gccatccgac ttggcctgtc tgcccgccat gtgcccagcc tcatcctgga gaccgagggc
1981 attccataca cactcaatgg caagaaagtg gaggtggccg tgaagcaggt gatggctggg
2041 aggactgtgg agcaccgggg ggcttctctc aaccccgaga ccctcgacct gtaccgggac
2101 atccctgagc tgcaggactt ctgagccagc agctcgact ccattccagct gtgcatgtga
2161 tggaaacttag ggacacttta gagacaacag ctgctccagg tggccctggc actgcacact
30 2221 ccacaggctc aggaacagct ttctctttc ggagtcattg gtgggggcca ggtctccga
2281 gctccaggat gtgcctggcc ttccggagac ctaagcatac actgaggctt gtctcgcccc
2341 tgctgctata ggtcagcaca cagccttgca ggggcagggc tggtgacatt tggtgacagc
2401 acactggagg agcgatgtgg ccttcagcct catgagcccc ctggtgtatc tcgctctatg
2461 aagggtgaagt cttacctata ctctctgtcc ttaactggcc tggcttctc agcgttactg
35 2521 tcttactcag agctgactaa agccagggca cctggcccag catgcctgtc acctcacagc
2581 tggctgttgg gacacctggg ctgacggctc ctgtaccgtc taccagggcc agaaccacat
2641 tctaggggaat ggtgagtga ttgtccagac agccaagagc tctctgctat ggcaagtccc
2701 ctctgctgtc ggggtgtccc gactgtttct gggctctacc agcacagtgt ttgaatgggt
2761 caaacggttt aatgttaatt gagggcctgg gggcacttga atccctgggt gtctaccca
40 2821 ggttaccact tagctgtgtg ctgggcagtg tgatcctgga gtccgccagg gatggagctg
2881 ccccgccctc gagggaaatg gccccgtgct ccctcctaag agttcctggg ctctttctat
2941 tcttactgtc ctctacattt gtaaatgaag tctcaggctt tggctgccag gacttaatgg
3001 tgcttctcct ctagcacacg gaatatgtca taaatcctgg tgattgcttt tgtatttttt
3061 tttttaaaaa ctgagcacia cataaaacct tttaaagata tcctggatct taagtctata
45 3121 aaggaaaagt gctatgaaga attttatgga ataaatctgt gccatgcaca cctgg

```

SEQ ID NO: 106

Amino acid sequence of mouse FLJ12389 encoded by the DNA sequence shown in SEQ ID NO: 105.

```

50 MSKLARLEREEIMBCQVMWEPDSKQDTQMDRFRAAVGTACGLALGNYNLDYHWSVRSYMD
FWAEFWKFSIGIVYSRMYDEVVDTSGKIADVPEWFRGSRNLNYAENLLRHKENDRVALYVAR
EGREEIVKVTFEELRQQVALFAAMRMKVKGDRVVGYLPNSAHAVEAMLAASIGAIW
SSTSPDFQVNGVLDLRFPSQIQPKLIFSVEAVVYNGKEHGHLEKLQRVVKGLPDLQRVVLI
55 YVLPREKIDISKIPNSVFLDDFLASGTGAQAPQLEFBLPFSHPLFIMFSSGTTGAPKCM
VHSAGGTLIQHLKEHMLHGNMTSSDILLYTTVGWMMWNMVSALATGASLVLYDGSPLV
PTFNVLWDLVDRIGITILGTGAKWLSVLEEKDMKPVETHNLHTLHTLSTGSPKQAQSYE
VYVRCIKSSVLLGSISGGTDIISCFMGQNSSIPVYKGEIQARNLGMAVEAWDEEGKAVWG

```

481 atcgtgaagg tgacttttga agagctgcgg cagcaggtgg ctctgttcgc agctgccatg  
 541 aggaagatgg gcgtgaagaa aggggaccgt gtggtcggct atctacccaa cagtgcgcac  
 601 gcgggtggagg ccatgctggc tgctgccagt attggggcca ttggagttc tacctcgccg  
 661 gactttggtg taaatggtgt cctggaccgc ttttctcaa ttcagccgaa acttatcttc  
 5 721 tcgggtggagg ctgttgtcta caatggcaag gagcatggtc acctggagaa gctgcagcga  
 781 gttgtgaaag gactgcctga cctgcagcga gtgggtgctga tccctatgt cctcccaagg  
 841 gagaagattg acatttccaa gatccccaac agtgtgttcc tggatgattt cctggcaagc  
 901 gggaccggcg cgcaggcgcc gcagctcgag tttagcagc tgccttcag ccacctctg  
 961 ttcacatcatgt tctcctcagg cagcagcggg gcgcccaggt gcatgggtga ctcagccggg  
 10 1021 ggcaccctca tccagcacct gaaggagcac atgctacacg gcaacatgac aagcagtgac  
 1081 atcctgctct actacaccac ggtcggctgg atgatgtgga actggatggg gtcagccctg  
 1141 gccacaggag catccttggg cctgtatgat ggctcccgcc tggttccgac ccccaacgtg  
 1201 ttgtgggacc ttgtggacag gatagggtac accatcctgg gaacgggagc caagtggctg  
 1261 tcagtgtctg aggagaagga catgaagcca gtggaaactc acaacctcca cagctgcac  
 15 1321 acgatcctgt ccaccggctc gccgtgaaa gccagagtt acgagtatgt gtacagatgc  
 1381 atcaagagct ccgtgctcct gggctccatc tcaggaggca ctgacatcat ctctgtttc  
 1441 atggggccaga actcctctat tctgtgtac aagggtgaga tccaagcccg gaaccttggc  
 1501 atggctgtgg aagcctggga cgaggaaggg aaggccgtct ggggagcgag tggcgagctg  
 1561 gtgtgcacta agcccatctc ctgagcacc acgcacttct ggaacgacga gaacggcagc  
 20 1621 aagtaccgga aggcctactt ctccaaattc ccagggtgtct gggcacacgg tgactactg  
 1681 aggatcaacc ccaaaacagg aggcattatc atgctggggc gtagtgatgg caccctcaac  
 1741 ccaatggcg tccgctttgg cagctcggag atctacaaca tcgtggaagc cttcgatgag  
 1801 gtggaggaca gcctgtgtgt accccagtac aacagagatg gcgaggagcg ggtggtcctg  
 1861 ttcctgaaga tggcgctccg gcacactttc cagcctgacc tcgtgaagcg catccgagac  
 25 1921 gccatccgac ttggcctgtc tgcccgccat gtgcccagcc tcactctgga gacccgagc  
 1981 attccataca cactcaatgg caagaaagtg gaggtggccg tgaagcagg gatggctggg  
 2041 aggactgtgg agcaccgggg ggcctctcc aaccccgaga ccctcgacct gtaccgggac  
 2101 atccctgagc tgcaggactt ctgagccagc agctcgact ccatccagct gtgcatgtga  
 2161 tggaaacttag ggacacttta gagacaacag ctgctccagg tggccctggc actgcacact  
 30 2221 ccacaggctc aggaacagct ttctctttc ggagtcattg gtgggggcca ggtctcccga  
 2281 gctccaggat gtgcctggcc ttcggagacc ctaagcatac actgaggctt gtctcgcccc  
 2341 tgctgctata ggtcagcaca cagccttgca ggggcagggc tggtgacatt tggtgacagc  
 2401 acactggagg agcgatgtgg ccttcagcct catgagcccg ctggtgtatc tcgctctatg  
 2461 aagggtgaagt cttacctata ctctctgtcc ttaactggcc tggcttctc agcgttactg  
 35 2521 tcttctcaag agctgactaa agccagggca cctggcccag catgctgtc acctcacagc  
 2581 tggctgttgg gacacctggg ctgacggctc ctgtaccgtc taccaggcc agaacaccat  
 2641 tctagggaat ggtgagtgaa ttgtccagac agccaagagc tctctgctat ggcaagtccc  
 2701 ctctgctgtc ggtgtgtccc gactgtttct ggtcttacc agcacagtgt ttgaatgggt  
 2761 caaacggttt aatgttaatt gagggcctgg gggcacttga atccctgggt gtctacccca  
 40 2821 ggttaccact tagctgtgtg ctgggcagtg tgatcctgga gtccgcccag gatggagctg  
 2881 ccccgccctca gagggaaaatg gcccgctgct ccctcctaag agttcctggg ctcttctat  
 2941 tgtgactgtc ctctacattt gtaaatgaag tctcaggctt tggctgccag gacttaatgg  
 3001 tgcttctcct ctgacacagc gaatatgtca taaatcctgg tgattgcttt tgtattttt  
 3061 tttttaaaaa ctgagcaca cataaaacct tttaaagata tcctggatct taagtctata  
 45 3121 aagggaaagt gctatgaaga attttatgga ataaatctgt gccatgcaca cctgg

SEQ ID NO: 106

Amino acid sequence of mouse FLJ12389 encoded by the DNA sequence shown in SEQ ID NO: 105.

50 MSKLARLEREEIMECQVMWEPDSKKDTQMDRFRAAVGTACGLALGNYNLDYHWSVRSYMD  
 FWAEFWKFSGIVYSRMYDEVVDTSGKIADVPEWFRGSRNLNYAENLLRHKENDRVALYVAR  
 EGREEIVKVTFEELRQQVALFAAAMRKMVGKGDVVGYPNSAHAVEAMLAASIGAIW  
 SSTSPDFGVNGVLDLRFQIQPKLIFSVEAVVYNGKEHGHLEKLQRVVKGLPDLQRVVLIP  
 YVLPREKIDISKIPNSVFLDDFLASGTGAQAPQLEFEQLPFSHPLFIMFSSGTTGAPKCM  
 55 VHSAGGTLIQHLKEHMLHGNMTSSDILLYTTVGWMMWNMVSALATGASLVLYDGSPLV  
 PTPNVLWDLVDRIGITILGTGAKWLSVLEEKDMKPVETHNLHTLHTILSTGSPKLAQSYE  
 VYVYRICSSVLLGSGTDSISCFMGNSSIPVYKGEIQARNLGMAVEAWDEEGKAVWG

ASGELVCTKPIPCQPTHFWNDENGSKYRKAYFSKFPVWAHGDYCRINPKTGGIIMLGRS  
 DGTLPNGVRFGSSEIYNIVEAFDEVEDSLCVPQYNRDGEERVVLFLKMASGHTFQPDLV  
 KRIRDAIRLGLSARHVPSLILETRGIPYTLNGKKVEAVKQVMAGRTVEHRGAFSPNPETL  
 DLYRDIPELQDF

5 SEQ ID NO: 107

gi|12831226|ref|NM\_023104.1| Rattus norvegicus acetoacetyl-CoA synthetase (LOC65984), mRNA

```

1 tctcgcgctg tggttcgtcc gcgccacgct gagccgctcc acgccttgcg ctctccgctg
61 tctccgcagc taaagcccg ggcagcccg ccacgcagct ccgcaaccat gtccaagctg
10 121 gcacggctcg agcgcgagga gatcatggag tgccagggtga tgtgggagcc tgacagcaag
181 aaggacacgc agatggaccg cttccggcg gccgtgggtg ctgctgcgg cctggcgctt
241 gggaattacg atgacttata ccactggtct gtcgggtcgt attcagactt ctgggctgag
301 ttctggaagt tcagtggagt tgtctgctct cgcagtgtatg atgaggttgt ggacacatcc
361 aaaggaattg cagatgtccc tgagtgggtc agaggcagcc gcctcaacta tgcagagaac
15 421 cttctgcggc acaaggagaa cgacagagtc gccctttacg tggcccgga aggcagagag
481 gagattgcga aggtgacttt cgaagagctt cggcagcagg tggctctgtt tgcagccgcc
541 atgaggaaga tgggcgtgaa gaaaggggac cgtgtggctg gttatctccc caacagtgcc
601 catgccgtgg aggccatgct ggctgtctgc agtattggag ccatttgagg ttctacctca
661 ccagactttg gtgtgaatgg tgtcctggac cgcttttctc aaattcagcc gaaacttatc
20 721 ttctcggtgg aagctgtggc ctacaacggc aaggaacacg gccacctgga gaagctgcag
781 cgagtcgtga aaggacttcc tgacctcag cgagtgggtc tgatccccta tgtcctccca
841 agggagaaga tagacatttc caagatcccc aacagcatgt ttctggatga cttcctggca
901 agcgggacag gtgcgcaggc accacagctc gaggttgaac agctgccctt cagccatccc
961 ctgttcatca tgttctcctc gggcacgaca ggagcgcca agtgcatggt gcactctgct
25 1021 gggggcaccc tcattccagc cctgaaggag cacgtgctac atggcaacat gacaagcagt
1081 gacatcctgc tctactacac caggtcgccg tggatgatgt ggaactggat ggtgtcagcg
1141 ctggccacag gacatcctt ggttctgtac gatggctccc cgctgggtcc aacacccaat
1201 gtgtgtggg accttgtgga caggatagga atcaccatcc tgggaacggg agccaagtgg
1261 ctgtcagtg caggagagaa ggacatgaag ccgatggaaa ctcacaacct ccacacgctc
30 1321 cacacgatcc tgtccaccgg ctccgactg aaagcccaga gctatgagta tgtgtacaga
1381 tgcatcaaga gcacgtgct cctcggtccc atctcagggt gcactgacat catctcctgt
1441 ttcattgggc agaactcatc tattcctgtg tacaagggtg agatccaagc ccggaacctc
1501 ggcatggccg tggaaagcctg ggacgaggaa gggaaaaccg tctggggagc gagtgcagag
1561 ctggtttgca ccaagcccat acctgcccag cccacgcact tctgggaacg cgagaacggc
35 1621 agcaagtaca ggaaggctta cttctccaaa taccagggtg tctgggcaca cggcgactac
1681 tgaggatca accccaagac aggaggatc gtcatgttg gccggagtga tggcaccctc
1741 aaccccaatg gcgtacgctt tggcagctcg gagatctaca acatttgga agccttcgat
1801 gaggtggagg acagcctttg tgtgcccag tacaacaggg atggtgagga gcggttagtc
1861 ctgtttctga agatggcctc tgggcacact ttccagcccg acctcgtgaa gcacatccgt
40 1921 gatgccatcc gccttggcct gtctgctcgc cacgtgccc aacctatcct ggagaccaa
1981 ggcattccat acacaatcaa cggcaagaaa gtggagggtg ccgtgaagca ggtgatagct
2041 ggggaagactg tggagcaccg gggggccttc tccaacctg agtccctgga cctgtatcgg
2101 gacatccctg agctgcagga cttctgaacc aatggctcgc actcggcca gctgtgcag
2161 tgatggaact tagggacagc agctgctccg gtggccagg cactacgcac tccacgggct
45 2221 cagagacagc ttccctgttt cggagtcag ggtggggacc aggtctgtct cctgagctcc
2281 aggatgtgcc tggccttcgg agacaaaag ggtacactga ggctgtctt ggccctgtct
2341 ctataggtca gcacacagc ttgcaggcgc aggtctggtg acatttggtg acaacacact
2401 ggaggagcga tgtggccttc agcctcatga gccacagtt ccacctcgt ctatgaagg
2461 gaaatcttac ctgtattctc tgcccttaat aactggtctg gcttctcag gtgtactgtc
50 2521 ttctcgagag ctgcctgaag ccagggcacc tgggtccagc ggctgtcac ctcacacttg
2581 gctgttgga tgcctgggct gatggctctg tctgccagg ccagaacacc attctaggga
2641 actgcgagt aatcctccag acagccaaga gctctctgct gtggcaagtc cctctgtctg
2701 gcgagcgtcc ccgactgttt ctgggtctta acagcacaag tatttgaatg ggtcaaactg
2761 caatgttact ggaggatctg ccgtctggag gcacttgaat cccctgggtg tctaccctag
55 2821 gttaccactt agctgtgtgc tgggcgggtg gatcctggag tcttccggg attgagactg
2881 ccggcctcag aggaaaatgg ccgcgtgtct cctcctaata gttccgggct ctttctcttg
2941 tgactgtact ccatgtttgt aaatgacgtc tcaggctttg gctgccagga ctcaatgggtg

```

3001 cttcttgtct agcacacaga atatgtcata aatcctgggtg attatacttt tgtattttctt  
 3061 tctttctttt ttttttttta aaactgagca caacataaaa ccttttaaaa atatcttgga  
 3121 ttttaactct ataagggaaa agtgctatga agaattttat ggaataaacc tgtgccatgc  
 3181 acgcctatcc

## 5 SEQ ID NO: 108

Amino acid sequence of rat FLJ12389 encoded by the DNA sequence shown in SEQ ID NO: 107.

MSKLARLEREEIMECQVMWEPDSKKDTQMDRFRAAVGTACGLALGNYDDLYHWSVRSYSYD  
 10 FWAEFWKPSGIVCSRMYDEVVDTSGKIADVPEWFRGSRNLNYAENLLRHKENDRVALYVAR  
 EGREEIAKVTFEELRQQVALFAAAMRKMGVKKGDRVVGYLPSAHAVEAMLAAASIGAIW  
 SSTSPDFGVNGVLDLRFSSQIQPKLIFSVEAVVYNGKEHGHLEKLQRVVKGLPDLQRVVLIP  
 YVLPREKIDISKIPNSMFLDDFLASGTGAQAPQLEFEQLPFSHPLFIMFSSGTTGAPKCM  
 VHSAGGTLIQHLKEHVLHGNMTSSDILLYTTVGWMMWNWMSALATGASLVLYDGSPLV  
 15 PTPNVLDLVDRLIGITILGTGAKWLSVLEEKDMKPMETHNLHTLHTILSTGSPKQAQSYE  
 YVYRCIKSTVLLGSISGGTDIISCFMGQNSSIPVYKGEIQARNLGMAVEAWDEEGKTVWG  
 ASGELVCTKPIPCQPTHFWNDENGSKYRKAYFSKYPGVWAHGDYCRINPKTGGIVMLGRS  
 DGTLPNGVRFSGSSEIYNIVEAFDEVEDSLCVPQYNRDGEERVVLFLKMASGHTFQPDLV  
 KHIRDAIRLGLSARHVPSLILETQGIPTYTINGKKVEAVKQVIAGKTVEHRGAFSPNESL  
 DLYRDIPELQDF

## 20 SEQ ID NO: 109

gi|22547160|ref|NM\_012193.2| Homo sapiens frizzled homolog 4 (Drosophila) (FZD4), mRNA

1 gctgcgcagc gctggctgct ggctggcctc gcggagacgc cgaacggagc cggccggcgc  
 61 cggcttgttg gctgcgcgcc tgcagccatg accctcgcag cctgtccctc ggccctcgcc  
 25 121 cgggacgtct aaaatcccac acagtgcgc gcagctgctg gagagccggc cgctgcccc  
 181 tcgtgcgccg atcacactcc cgtcccggga gctgggagca gcgcgggcag ccggcgcccc  
 241 cgtgcaaaact ggggggtgtct gccagagcag cccagccgc tgcgcgtgct acccccgatg  
 301 ctggccatgg cctggcgggg cgagggccg agcgtcccgg gggcgccgg gggcgctcgg  
 361 ctgagtcagg ggtgtctcct gcagttgctg ctgctcctgg ggcggcgcg gggcttcggg  
 30 421 gacgaggaag agcggcgctg cgacccatc cgcattccca tgtgccagaa cctcggctac  
 481 aacgtgacca agatgcccac cctgggtggg cacgagctgc agacggagc cgagctgcag  
 541 ctgacaactt tcacaccgct catccagtac ggctgctcca gccagctgca gttcttcctt  
 601 tgttctgttt atgtgccaat gtgcacagag aagatcaaca tccccattgg cccatgcggc  
 661 ggcatgtgtc tttcagtcaa gagacgtgt gaacccgtcc tgaaggaatt tggatttgcc  
 35 721 tggccagaga gtctgaactg cagcaaattc ccaccacaga acgaccacaa ccacatgtgc  
 781 atggaagggc caggtgatga agaggtgccc ttacctaca aaaccccat ccagcctggg  
 841 gaagagtgtc actctgtggg aaccaattct gatcagtaca tctgggtgaa aaggagcctg  
 901 aactgtgtgc tcaagtgtgg ctatgatgtt ggcttataca gccgctcagc caaggagtgc  
 961 actgatattt ggatggctgt gtgggccagc ctgtgtttca tctccactgc cttcacagta  
 40 1021 ctgaccttcc tgatcgattc ttctagggtt tcctaccctg agcggcccat catatttctc  
 1081 agtatgtgct ataataattt tagcattgct tatattgtca ggctgactgt aggcggggaa  
 1141 aggatatact gtgattttga agaggcagca gaacctgttc tcatccaaga aggacttaag  
 1201 aacacaggat gtgcaataat ttcttctgt atgtactttt ttggaatggc cagctccatt  
 1261 tgggtgggtta ttctgacact cacttggttt ttggcagcag gactcaaata gggcatgaa  
 45 1321 gccattgaaa tgcacagctc ttatttccac attgcagcct gggccatccc cgcagtgaaa  
 1381 accattgtca tcttgattat gagactgggt gatgcagatg aactgactgg cttgtgctat  
 1441 gttggaaacc aaaatctcga tgccctcacc ggggtcgtgg tggtccccct ctttacttat  
 1501 ttgggtcattg gaactttgtt cattgtgtgc ggtttgggtg cttgtttcaa aattcgggtc  
 1561 aatcttcaaa aggatgggac aaagacagac aagttagaaa gactgatggt caagattggg  
 50 1621 gtgttctcag tactgtacac agttcctgca acgtgtgtga ttgcctgtta ttttatgaa  
 1681 atctccaact gggcactttt tcggtattct gcagatgatt ccaacatggc tgttgaaatg  
 1741 ttgaaaattt ttatgtcttt gttgggtggc atcacttcag gcatgtggat ttggtctgcc



1801 aaaactcttc acacgtggca gaagtgttcc aacagattgg tgaattctgg aaaggtaaag  
 1861 agagagaaga gaggaaatgg ttgggtgaag cctggaaaag gcagtgaagac tgtggtataa  
 1921 ggctagtcag cctccatgct ttcttcattt tgaagggggg aatgccagca ttttggagga  
 1981 aattctacta aaagttttat gcagtgaatc tcagtttgaa caaactagca acaattaagt  
 5 2041 gaccccgctc aacccactgc ctcccacccc gaccccgagca tcaaaaaacc aatgattttg  
 2101 ctgcagactt tggaaatgat caaaatggaa aagccagtta gaggctttca aagctgtgaa  
 2161 aaatcaaaac gttgatcact ttagcaggtt gcagcttgga gcgtggagggt cctgcctaga  
 2221 ttccaggaag tccagggcga tactgttttc cctgcagggg tgggatttga gctgtgagtt  
 2281 ggtaactagc agggagaaat attaaacttt ttaacccttt accattttta atactaactg  
 10 2341 ggtcttttcag atagcaaagc aatctataaa cactggaaac gctgggttca gaaaagtgtt  
 2401 acaagagttt tatagtttgg ctgatgtaac ataaacatct tctgtggtgc gctgtctgct  
 2461 gtttagaact ttgtggactg cactcccaag aagtgggtgtt agaactcttc agtgcctttg  
 2521 tcataaaaca gttatttgaa caaacaaaag tactgtactc acacacataa ggatccaggt  
 2581 ggattttttct tctctgtcct cctctcttaa atttcaacat ctctcttctt ggctgctgct  
 15 2641 gttttcttca ttttatgtta atgactcaaa aaagggtattt ttatagaatt tttgtactgc  
 2701 agcatgctta aagaggggaa aaggaagggt gattcacttt ctgacaatca ctttaattcag  
 2761 aggaaaatga gatttactaa gttgacttac ctgacggacc ccagagacct attgcattga  
 2821 gcagtgggga cttaatatat tttacttgtg tgattgcac tcagcagacg ccagtctgga  
 2881 agagctgaaa tgttaagttt cttggcaact ttgcattcac acagattagc tgtgtaattt  
 20 2941 ttgtgtgtca attacaatta aaagcacttt gttggaccat gacatagat actcaactga  
 3001 ctttaaaact atgggtcaact tcaacttgca ttctcagaat gatagtgctt ttaaaaattt  
 3061 ttttattttt taaagcataa gaatgttatt agaactctgt ctacttagga caatggagac  
 3121 tttttcagtt ttataaaggg aactgaggac agctaattcca actacttggt gcgtaattgt  
 3181 ttcttagtaa ttggcaaaag ctctctgtta gatttcactg gaggcagtggt ggcttgaggt  
 25 3241 atttatatgg tgcttaatga atctccagaa tgccagccag aagcctgatt ggttagtagg  
 3301 gaataaagtg tagaccatat gaaatgaact gcaaaactca atagcccagg tcttaattgc  
 3361 ctcttagcaga ggtatccaaa gcttttaaaa ttatgcata cgttcttcac aaggggggtac  
 3421 cccagcagc ctctcgaaaa ttgcacttct cttaaaactg taactggcct ttctcttacc  
 30 3481 ttgccttagg ccttctaatt atgagatctt ggggacaaat tgactatgtc acagggtgct  
 3541 ctctctgtta ctcatacctg tctgtctcag caactgcttt gcaatgacat ttatttatta  
 3601 attcatgcct taaaaaaata ggaaggggag cttttttttt tctttttttt ttttcaatc  
 3661 acacttttggt gaaaaacatt tccagggact caaaattcca aaaagggtgtt caaattctgtg  
 3721 aagtaagcat ttctctttt tttaaaattt ggtttgagcc ttatgcccac agtttgacat  
 3781 ttctctttct tcttctttt ttgtttttt gtggttctga agctctctga catcaagatg  
 35 3841 catgtaaagt cgattgtatg ttttggaagt caaagtcctg gcttttgaga ctgaggttaa  
 3901 gtgggcacag gtggccctgt ctgctgtgcc cagtctgagt accttggtta gactctaggt  
 3961 caggctocag gagcatgaga attgatcccc agaagaacca ttttaactcc atctgatact  
 4021 ccattgccta tgaaatgtaa aatgtgaact ccctgtgctg cttgtagaca gttcccataa  
 4081 ctgtccacgg ccctggagca cgcacccagg ggcagagcct gcccttactc acgctctgct  
 40 4141 ctggtgtctt gggagtgtgt cagggaactc ggcccaggca ggggaaggaa gaccaggcgg  
 4201 taggggactg gtctgtgtgt tagagtatag aggtttgtta tgcagttttc ttcataatgt  
 4261 gtcagtgatt gtgtgacaa ggcagcatct agcagaaagc caggcatgga gtagtgatc  
 4321 gatacttgct aatgactaaa taataacaat aaaagagcac ttgggtgaat ctgggcacct  
 4381 gatttctgag ttttgagttc tggagctagt gttttgacaa tgctttgggt tttgacatgc  
 45 4441 cttttccaca aatctcttgc cttttcaggg caaagtgtat ttgatcagaa gtggccattt  
 4501 ggattagtag ccttagcaat gctacaggggt tataggcctc tcctttcaca ttccagacaa  
 4561 tggagagtgt ttatggtttc aggaaaagaa ctttgtgggt gaggggtcag ttaccagtga  
 4621 ccttcaatca actccatcac ttcttaaatc ggtatttgtt aaaaaaatca gttattttat  
 50 4681 ttattgagtg ccgactgtag taaagccctg aaatagataa tctctgttct tctaactgat  
 4741 ctaggatggg gacgcaccca ggtctgtctga actttactgt tctctggga aaggagcagg  
 4801 gacctctgga attcccatct gtttactgt ctccattcca taaatctctt cctgtgtgag  
 4861 ccaccacacc cagcctgggt ctctctactt ttaacacatc tctcatccct ttcccaggat  
 4921 tccttccaag tcagttacag gtggttttaa cagaagcat cagctctgct tcgtgacagt  
 4981 ctctggagaa atcccttagg aagactatga gagtaggcca caaggacatg ggcccacaca  
 55 5041 tctgctttgg ctttgccggc aattcagggc ttgggggtatt ccattgtgact tgtataggta  
 5101 tatttgagga cagcatcttg cttagaataa ggtgaggggt gtttttcttt gtttgaaac  
 5161 tacagtaaat gggatgatt gtatgttctt cagaaatccc ttggcctcca gagatataac  
 5221 atgggtgcaat ggcacctctg tccaacctcc tttctggtag attcctttct cctgcttcat  
 5281 ataggccaaa cctcagggca agggaacatg ggggtagagt ggtgctggcc agaaccatct  
 60 5341 gcttgagcta cttgggtgat tcatactctc tttcctttat ggagacccat ttctgtatct  
 5401 ctgagactgt tgctgaactg gcaacttact tgggcctgaa actggagaag ggggtgacatt

```

5461 tttttaattt cagagatgct ttctgatttt cctctcccag gtcactgtct cacctgcact
5521 ctccaaactc aggttccggg aagcttggtt gtctagatac tgaattgaga ttctgttcag
5581 caccttttag ctctatactc tctggctccc ctcatcctca tggctactga attaaatgct
5641 tattgtattg agaaccaaga tgggacctga ggacacaaag atgagctcaa cagtctcagc
5 5701 cctagaggaa tagactcagg gatttcacca ggtcgggtgca gtatttgatt tctggtgagg
5761 tgaccacagc tgcagttagg gaagggagcc attgagcaca gactttggaa ggaacctttt
5821 ttttggtgtt tgtttgtttg tttgtttgtt tgtttgtttg agacagggtc ttgctctgtc
5881 acccaggctg gggcgcaatg gcacgatctt ggctcactgc aacctctgcc tctggggttc
5941 aagtgattct cctgccacag cctcctgagg agctgggact acaggtgcbt gctaccacgc
10 6001 ccagctactt ctgtattttt agtagagacg ggggtttcact gtgttgggca ggctggcttc
6061 gaactcctga cctcatgatc tgcccgcctc agcctcccaa agtgctggga ttacaagtgt
6121 gagccaccac acctggcctg gaaggaacct cttaaaatca gtttacgtct tgtatttgtt
6181 tctgtgatgg aggacactgg agagagttgc tattccagtc aatcatgtcg agtcaactgga
6241 ctctgaaaat cctattgggt cctttatttt atttgagttt agagttccct tctggggttg
15 6301 tattatgtct ggcaaatgac ctgggttatt acttttcctc cagggttaga tcatagatct
6361 tggaaactcc ttagagagca ttttgctcct accaaggatc agatactgga gccccacata
6421 atagatttca tttcactcta gcctacatag agctttctgt tgctgtctct tgccatgcac
6481 ttgtgcggtg attacacact tgacagtacc aggagacaaa tgacttacag atcccccgac
6541 atgcctcttc ccttgccaa gctcagttgc cctgatagta gcatgtttct gttctgtatg
20 6601 tacctttttt ctcttctctt ttgcatcagc caattcccag aatttcccca ggcaatttgt
6661 agaggacctt tttggggtcc tatatgagcc atgtcctcaa agcttttaaa cctccttgct
6721 ctctacaat attcagtaca tgaccactgt catcctagaa ggcttctgaa aagaggggca
6781 agagccactc tgcgccacaa aggttgggtc catcttctct cagaggttgt gaaagttttc
6841 aaattgtact aataggctgg ggccctgact tggctgtggg ctttgggagg ggtaagctgc
25 6901 tttctagatc tctcccagtg aggcattggg gtgtttctga attttgccta cctcacaggg
6961 atgttggtgag gcttgaaaag gtcaaaaaat gatggccctt tgagctcttt gtaagaaagg
7021 tagatgaaat atcggtatga atctgaaaaa aagataaaat gtgacttccc ctgctctgtg
7081 cagcagtcgg gctggatgct ctgtggcctt tcttgggtcc tcatgccacc ccacagctcc
7141 aggaaccttg aagccaatct gggggacttt cagatgtttg acaaagaggt accaggcaaa
30 7201 cttcctgcta cacatgcctt gaatgaattg ctaaatttca aaggaaatgg accctgcttt
7261 taaggatgta caaaagtatg tctgcatcga tgtctgtact gtaaatttct aatttatcac
7321 tgtacaaaga aaacccttg ctatttaatt ttgtattaaa ggaaaataaa gttttgtttg
7381 ttaaaaaaaa a

```

SEQ ID NO: 110

35 Amino acid sequence of human FZD4 encoded by the DNA sequence shown in SEQ ID NO: 109.

```

MAWRGAGPSVPGAPGGVGLSLGLLQLLLLLLGPARGFGDEEERRCDPIRISMCONLQYNV
TKMPNLVGHELQTDALQLTFTPLIQYGCSSQLQFFLCVYVPMCTEKINIPIGPCGGM
CLSVKRRCEPVLKEFGFAWPESLNCSEKFPQNDHNHMCMEGPGDEEVPLPHKTPIQPGEE
40 CHSVGTNSDQYIWKRLNLCVLCGYDAGLYSRSAKEFTDIWMAVWASLCFISTAFVLT
FLIDSSRFSYPERPIIFLSMCYNIYSIAYIVRLTVGRERISCFEAAEPVLIQGLKNT
GCAIIFLLMYFFGMASSIWWVILTLTWFLAAGLKWGHEAIEHSSYFHIAAWAIPAVKTI
VILIMRLVDADELTLGLCYVGNQNLDAITGFVVAFLFTYLVIGTLFIAAGLVALFKIRSNL
QKDGTKTKDLERLMVKIGVFSVLYTVPATCVIACYFYEISNWALFRYSADDSNMAVEMLK
45 IFMSLLVGITSGMWIWSAKTLHTWQKCSNRLVNSGKVKREKRGNGWVKPGKSETVV

```

SEQ ID NO: 111

gi|13548680|dbj|AB054881.1| Homo sapiens mRNA for Soluble-type polypeptide FZD4S, complete cds

```

50 1 atccacacac gtcgcgcgca gctgctggag agccggccgc tgccccctcg tcgccgcac
61 aactccccgt cccgggagct gggagcagcg cgggcagccg gcgccccctg gcaaactggg
121 ggtgtctgcc agagcagccc cagccgctgc cgctgctacc cccgatgctg gccatggcct
181 ggcgggggcg agggccgagc gtcccggggg cgccgggggg cgctgggtctc agtctggggg

```

241 tgctcctgca gttgctgctg ctctctggggc cggcgcgggg cttcggggac gaggaagagc  
 301 ggcgctgcca ccccatccgc atctccatgt gccagaacct cggctacaac gtgaccaaga  
 361 tgcccaacct gggtggggcac gagctgcaga cggacgccga gctgcagctg acaactttca  
 421 caccgctcat ccagtagggc tgctccagcc agctgcaggt gggcgccccc acccccaccc  
 5 481 ctggcgggac cccttgggga gggacgctcc aaactaactt tgcagagccc atgccaagtt  
 541 gatcatccta gcccgaaagg gacagctctc ctggaaaagt gatttccatc cccaccccca  
 601 ctttctcagg gactggaagc caaaacgttg tgtaagtcac ctggcctggg aaagaaccca  
 661 ctcttacacc ccgccccttc gtttttctct cttgcccacc catgtctggc ccagaccctt  
 721 aacccagtg gagctgaggg ttatctctgc caaggattct ggccgccact gctcgggtggg  
 10 781 cgagtgggag ggagccagct gcagctaaga cttggaggga gtgataagga gaccatcctt  
 841 caccgaactt ccttctcccc actatttttg ggggtgttatt aaatgaaaca ctactgcacc  
 901 ttttctgct gaatagcccc ttctcacgac gtcccgcagc gttttagagg tcatgcatga  
 961 agagtgggt ggggtggctt gaggttcttc ttatctcac agttgaaggc acgttttaag  
 1021 cttggagggt gagaagaagg tgcaggagg tggttggtat attggaagaa atttttttgc  
 15 1081 agtcattaaa aaatgtaaag catatctaag gatagagaaa atgtttattc cacagtaata  
 1141 agaatttgca tatacagggt gattataatc ctgcaaaaata aaaaatttat tgggataata  
 1201 aaagactgac aggaaaattt ttaaagtgtt aacattggtt atgtttgggt tggtagggt  
 1261 gggtagattt tagaaattta gaaaagagag aagttgtggg aaaaagtaga cgtaggtgt  
 1321 ccaggcttca gaaatctggt ttctggccgg gcgcagtggc tcacgcctgt aatcccagca  
 20 1381 ctttggagg cggaggcagg tggctcacga ggttaggag tcgagaccag cctgaccaac  
 1441 atggcgaaat cctgtctcta ctaaaagtac aaaaatttag cgggcgtggt ggcgtgcgcc  
 1501 tgtaatccca gctactcagg aggtcagggc aggagaatcg cttgaaccca ggaggcagag  
 1561 gttgcagtg ggcgggattg caccattgca ctctagcctg ggccacagga gcgaaactcc  
 1621 gtatcaaaaa aaaaaagaaa agaaaagaaa tctgtggttc tggctgtaac ttctgagaga  
 25 1681 taataaacag acatacagct tacttactag gtgcatgttt ttacataggt cattttaaaa  
 1741 cttggtttgt ggttttgga agatggtggg gattagacct gttgctgcta aaatatccct  
 1801 ttggaagtgt cagaacttta gaactaggtt ataaaagctt aagaaatgat cctccagctc  
 1861 tcagtttggg aacagcacat atctgacat cagtggcaat ttgttgga aacagctttt  
 1921 gcagggatata atatttttaa ttacatgtat cctggggaaa gcggccaagc tgttttgaag  
 30 1981 gacaagactg gatcccttta catgctggaa aatagttact tgtaactcta gacttcatag  
 2041 acaacatctg taaggagcca agcagactgt ttatcaatgc tggagggaatt gaggacagca  
 2101 actggactct cccctgctgg tactgcagtt tcagcgactt aacatttaca ttttcctatt  
 2161 gtgcttcatt attattcatt gaggaacctt ggccaaacaa atccctagt accagagacc  
 2221 tcaatctgca tccatccctt gactctaaca gtgtctcatt tgggccattt tttgtttcag  
 35 2281 tatgggacta gtttaaacaa atttttaagg gataaattag ggagaagaga atactaacag  
 2341 gatctacatt tttttttaac ttttattttt aaaaatgataa agaaacgggg ataaaaataa  
 2401 accacctaata tgagtcatat tagtttcttt agctttttca aatagtttct aaaatggaag  
 2461 gagattgaaa gccagttttt taagatttat ttcaagtgtt gtgcaaagag ggcttatgca  
 2521 attcttcaga ccctatttat tttatttatt tatttatgta cattctctcg gaaaggatta  
 40 2581 ctaaagccaa actaagaaaa gtatgtaagt gatggattat caggcgccct tgtttgaggt  
 2641 cagagtaaac agacagactc ccgataggtt ggtttaaaag tagataccta gattgcctgg  
 2701 aagcattcaa ctacagctttg tgggagcatt tggctcaaat tccaagtcac tttttttt  
 2761 gtgtgtgttc attttgttac tttccagtt ctctctttgt tctgtttatg tgccaatgtg  
 2821 cacagagaag atcaacatcc ccattggccc atgcggcggc atgtgtcttt cagtcaagag  
 45 2881 acgctgtgaa cccgtctga aggaatttgg atttgccctg ccagagagtc tgaactgcag  
 2941 caaattccca ccacagaacg accacaacca catgtgcag gaagggccag gtgatgaaga  
 3001 ggtgccctta cctcacaaaa ccccatcca gcctggggaa gagtgtcact ctgtgggaac  
 3061 caattctgat cagtacatct ggtgaaaag gagcctgaac tgtgtgctca agtgtggcta  
 3121 tgattctggc ttatacagcc gctcagccaa ggagttcact gatctctgga tggctgtgtg  
 50 3181 ggccagcctg tgtttcatct ccactgcctt cacagtactg accttctctga tctgattcttc  
 3241 taggttttcc taccctgagc gcccacatcat atttctcagt atgtgtctata atatttatag  
 3301 cattgcttat attgtcaggg tgactgtagg ccgggaaagg atatcctgtg attttgaaga  
 3361 ggcagcagaa cctgttctca tccaagaagg acttaagaac acaggatgtg caataatttt  
 3421 cttgctgatg tacttttttg gaatggccag ctccatttgg tgggttatcc tgacactcac  
 55 3481 ttggtttttg gcagcaggac tcaaatgggg tcatgaagcc attgaaatgc acagctctta  
 3541 tttccacatt gcagcctggg ccacccccgc agtgaaaacc attgtcatct tgattatgag  
 3601 actgggtggat gcagatgaac tgactggctt gtgctatgtt ggaaaccaa atctcgatgc  
 3661 cctcacccggg ttctgtgttg ctccccctct tacttatttg gtcattggaa cttgtttcat  
 3721 tgctgcaggt ttggtggcct tgttcaaaat tgggtcaaat cttcaaaagg atgggacaaa  
 60 3781 gacagacaag ttagaaagac tgatgggtcaa gattgggggtg ttctcagtac tgtacacagt  
 3841 tcctgcaacg tgtgtgattg cctgttattt ttatgaaatc tccaactggg cactttttcg

```

3901 gtattctgca gatgattcca acatggctgt tgaaatgttg aaaattttta tgtctttggt
3961 ggtgggcatc acttcaggca tgtggatttg gtctgccaaa actcttcaca cgtggcagaa
4021 gtgttccaac agattggtga attctggaaa ggtaaagaga gagaagagag gaaatggttg
4081 ggtgaagcct ggaaaaggca gtgagactgt ggtataaggc tagtcagcct ccattgcttc
5 4141 ttcattttga aggggggaat gccagcattt tggaggaaat tctactaaaa gttttatgca
4201 gtgaatctca gtttgaacaa actagcaaca attaatgtac ccccgtaac ccactgcctc
4261 ccaccccgac cccagcatca aaaaaccaat gatcttgctg cagactttgg aatgatccaa
4321 aatggaaaag ccagtttagag gctttcaaaag ctgtgaaaaa tcaaacggt gatcacttta
4381 gcagggttga gcttgagcgt tggaggtcct gcctagattc caggaagtcc agggcgatac
10 4441 tgttttcccc tgcagggtgg gatttgagct gtga

```

SEQ ID NO: 112

Amino acid sequence of human FZD4 variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 111.

```

15 MLAMAWRGAGPSVPGAPGGVGLSLGLLLQLLLLGPARGFGDEEERRCDPIRISMCONLG
YNVTKMPNLVGHELOTDAELQLTFTPLIQYGCSSQLQVGAPTPTPGGTPWGGTLQTNFA
EPMPS

```

SEQ ID NO: 113

gi|15929644|gb|BC015256.1| Mus musculus frizzled homolog 4 (Drosophila), mRNA (cDNA clone MGC:18403 IMAGE:4238940), complete cds

```

20 1 agcgctgggg cggtgagaac agcgcggcgt agagtgcagg cgggcttcgc cgaaaagccg
61 gactcggccg gcgccgagtt ctgggatcgc cgctgcagc catgacccta gcagtcctac
121 cctcggcccc ggctccggac gtctgatata ccgcacattc tcgtacaact gctggagagg
181 cgactgctgc ccccttgctg cccttggcgc cttaccgcat tccctatccg gagttgggag
241 cagcgcggcc accggcgccc ctgtgcaaac tgggggtgtc tgctagatca gcctctgccg
25 301 ctgtgcccgc cagctctggc catggcctgg ccgggcacag ggccgagcag ccggggggcg
361 cctggaggcg tcgggctcag gctggggtgt ctgtgcagc tgctcctgct cctgcggccg
421 acattggggg tcggggacga ggaggagcgg cgctgcgacc ccacccgcat cgccatgtgc
481 cagaacctcg gctacaacgt gaccaaagat cccaacttag tgggacacga gctgcagaca
541 gcgcccagc tgacgtgac aactttcacg ccgctcatcc agtacggctg ctccagccag
30 601 ctgcagttct tcctttgttc ggtttatgtg ccaatgtgca cagagaagat caacatcccc
661 atcgcccgct gcggtggcat gtgcctttca gtcaagagac gctgtgaacc agtcttgaga
721 gaatttgggt ttgcctggcc cgacacctg aactgcagca agtcccgcc ccagaacgac
781 cacaaccaca tgtgcatgga aggaccaggt gatgaagagg ttcccttgcc ccacaagact
841 cccatccagc ccggggaaga gtgccactcc gtgggaagca attctgatca gtacatctgg
35 901 gtgaagagga gcctgaactg tgtctcaag tgtggctacg atgtggctt gtacagccgc
961 tcagctaagg agttcacgga tatttggatg gctgtgtggg ccagcctctg cttcatctcc
1021 accaccttca ccgtgctgac ctctctgatt gattcatcca ggttttctta ccttgagcgc
1081 cccatcataat ttctcagtat gtgctataat atttatagca ttgcttataat tgttcggctg
1141 actgtaggcc gggaaaggat atcctgtgat tttgaagagg cggcagagcc cgttctcatc
40 1201 caagaaggac ttaagaacac aggatgtgca ataattttct tgctgatgta cttttttgga
1261 atggccagct ccatttggtg ggttattctg aactcactt ggtttttggc agccggactc
1321 aagtggggtc atgaagccat tgaaatgcac agttcttatt tccacatcgc agcctgggct
1381 attcccgtag tgaaaacat tgctatcttg attatgagac tagtggatgc cgtgaactg
1441 actggcttgt gctatgttgg gaacaaaaac ctagatgccc tcaactggctt tgtgtggctt
45 1501 cctctcttta cgtatttggg gattggaacg ctgttcattg cggcggggtt ggtggcctta
1561 ttcaaaattc ggtccaatct tcaaaaagac gggacaaaga cagacaagtt ggaaggcta
1621 atggtcaaga tcggggtctt ctcagtagtg tacacgggtt ctgcaacctg tgtgattgcc
1681 tgttatttct atgaaatctc aaactgggca ctctttcgat attctgcaga tgactcaaac
1741 atggcagttg aaatgttgaa aatttttatg tctttgctcg tgggcatac ttcaggcatg
50 1801 tggatttggg ctgcaaaaac tcttcacacg tggcaaaagt gttctaaccg attggtgaat
1861 tctgggaagg taaagagaga gaagaggggg aatggttggg tgaagccagg aaaaaggcaac
1921 gagactgtgg tataagacta gccggcttcc tcgttcctca ttgtgaagga agtgatgcag

```

1981 ggaatctcag tttgaacaaa cttagaacaa cttcagccca cacacaccca cgtcagccca  
 2041 ccaccactca cccaactcag catcagaaga ccaatggctt cactgcagac tttggaatgg  
 2101 tccaaaatgg aaaagccagt tagaggtttt caaagctgtg aaaaatcaaa atgttgatca  
 2161 ctttagcagg tcacagcttg gagtcctggt aggtcccgc tagattcctg aagcccaggg  
 5 2221 tgatagtgtt tgctcctact ggggtgggatt tcaactgtga gttgataaca tgcaaggaga  
 2281 aagattaatt tttaaaaccc ttttaaattt taaatagtaa ctaggtcttg cagatagcaa  
 2341 agtgatctat aaacactgga aatgctgggt tgggagacgt gttgcagagt tttatagttt  
 2401 ggctggctta acataaacat cttctggcct acactgtctg ctgttttagaa ctctgtagcg  
 2461 cactcccaag aggtgggtgtc aaaatccttc agtgcctttg tcgtaaaaca gaattggttg  
 10 2521 agcaaacaaa agtactgtac taacacacgt aaggatccca gtggatttct ctctcctgaa  
 2581 atttcaacat ccctaattct aggcagcccc tgttttcttc actttaaact aatgactcaa  
 2641 aaaaaaaaag gttattttta taggattttt ttttgcaact cagcatgcct aatgagagga  
 2701 aaaggggaag tgattcactt tctgacaatc acttaattca gagaaaaatg agcaattgcta  
 2761 agttgactta ccttaccgac cctagagacc tattgcatta agcaatgtta agcaattggg  
 15 2821 acttaaaata ttttagtttg tgtgattgca tctaggcaga cgccagctctg gaagaactga  
 2881 aatgttaaat ttcttggcaa ctttgcattc acacagatta actgtgtaat ttgtgtgtgt  
 2941 caattacaat taaaagcaca ttcttggacc atgaaaaaaa aaaaaaaaaa

SEQ ID NO: 114

20 Amino acid sequence of mouse FZD4 encoded by the DNA sequence shown in SEQ ID NO: 113.

MAWPGTGPSSRGAPGGVGLRLGLLLQLLLLRLPTLFGDEEERRCDPIRIAMCONLGYNV  
 TKMPNLVGHELQTDALQLTFTPLIQYGCSSQLQFFLCVYVPMCTEKINIPIGPCGGM  
 CLSVKRRCEPVLREFGFAPDNLNCSKFPFQNDHNMCMEGPGDEEVLPHKTPIQGEE  
 CHSVGSNSDQYIWKRLNLCVLCGYDAGLYSRSAKEFTDIWMAVWASLCFISTTFTVLT  
 25 FLIDSSRFSPYPERPIIFLSMCYNIYSIAYIVRLTVGRERISCDFEEAAEPVLIQEGLKNT  
 GCAIIFLLMYFFGMASSIWWVILTLTWFLAAGLKWGHEAIEMHSSYPHIAAWAIPAVKTI  
 VILIMRLVDADELTLGLCYVGNQNLDAITGFVVAFLFTYLVIGTLFIAAGLVALFKIRSNL  
 QKDGTKTKLERLMVKIGVFSVLYTVPATCVIACYFYEISNWALFRYSADDSNMAVEMLK  
 IFMSLLVGITSGMWIWSAKTLHTWQKCSNRLVNSGKVKREKRGNGWVKPGKNETVV

30 SEQ ID NO: 115

gi|12018309|ref|NM\_022623.1| Rattus norvegicus frizzled homolog 4 (Drosophila) (Fzd4), mRNA

1 gagcttggtta actagagcag gctcattcat taaatcgtgc acatgattct acagcagctg  
 61 aagaagcggc cgctgcccc tgcgcgcctt tggcgcccta ccgcattccc tatccggagc  
 35 121 tgggagctgc gcggccaccg gcgcacctgt gcaaactggg ggtgtctgct agagcagcct  
 181 ccgcgcgtgc tgcccgcggc tctggccatg gcttggcagg gcacaggggc aagcgtccgg  
 241 gggatgcctg gaggcgtcag gctcaggctg gggctgctgc tgcctgcagt gctcctgctc  
 301 cagcggcccc ccctgggggt cggggacgag gaggagcggc gttgcgacct catccgcatc  
 361 gccatgtgcc agaaccctcg ctacaacgtg accaagatgc ccaacttagt gggacacgag  
 40 421 ctgcagacag acgcggagct gcagctgaca actttcacgc cgctcatcca gtacggctgc  
 481 tccagccagc tgcagttctt cctttgttcg gtttatgtgc caatgtgcac agagaagatc  
 541 aacatcccca tcggcccgtg cgggtggcatg tgcctttcag tcaagagacg atgtgaacca  
 601 gtcctgaaag aatttggtt tgccctggcc gacagcctga actgcagcaa gttcccaccc  
 661 cagaacgacc acaaccacat gtgcattgaa ggaccagggt acgaagaggt acccttgccc  
 45 721 cacaaaactc ccatccagcc gggggaagag tgccactccg tgggaaccaa ttccgacag  
 781 tacatctggg tgaaaaggag cctgaactgt gttctcaagt gtggctacga tgctggcttg  
 841 tacagccgct cagctaagga gttcacggat atttgatgg ccgtgtgggc cagcctctgc  
 901 ttcattctcta ccaccttcac tctgctgacc ttcttgattg attcgtccag gttttcttac  
 961 cctgagcgcc ccatcatatt cctcagtatg tgctataata tttatagcat tgcttatatt  
 50 1021 gttcggctta ctgtaggccg ggaaaggata tcctgtgatt ttgaagaggc ggcagaaccc  
 1081 gttctctacc aagagggact taagaacaca gcatgtgcaa taattttctt gctgatgtac  
 1141 ttttttgtaa tggccagctc catttgggtg gttattctga cactcacttg gtttttgga

```

1201 gccggactca agtgggggtca cgaagccatt gaaatgcaca gttcttattt ccacatcgca
1261 gcctgggcta tccctgccgt gaaaaccatt gtcattctga ttatgagact agtggatgcc
1321 gatgagctga ctggcctgtg ctatgtcggg aaccaaagcc tagatgccct caccggcttt
1381 gtggtggcac ctctctttac ctatttggtg attggaactc tattcattgc tgcaggcttg
5 1441 gtggccctct tcaaaattcg gtcaaactct caaaaagatg ggaccaagac agacaagttg
1501 gaaaggctaa tggtaagat cgggggtctc tcggtcctgt acacggttcc tgcgacctgt
1561 gtgattgcct gttatttcta tgaaatctca aactgggcac tctttcggta ttctgcagat
1621 gactcaaaca tggcagttga aatggtgaaa atttttatgt ctttgcctgt gggcatcact
1681 tcaggcatgt ggatttggtc tgccaaaact cttcacacgt ggcaaaaatg ttctaaccga
10 1741 ttggtgaatt ctgggaaggt aaagagagag aagaggggga atgggtgggt gaagccaggg
1801 aaaggcaacg aaactgtggt gtaggactag cttagcttcca ctttctcat tttgaaggaa
1861 aggatgcagt gaatctcagt ttgaacaaac tagaaaca

```

SEQ ID NO: 116

Amino acid sequence of rat FZD4 encoded by the DNA sequence shown in SEQ ID NO: 115.

```

15 MAWQGTGPSVRGMPGGVRLRLGLLLQLLLLQRLPALGFGDEEERRCDPIRIAMCQNLGYN
VTKMPNVLGHEQLTDAELQLTTFTPLIQYGCSSQLQFFLCVYVPMCTEKINIPIGPCGG
MCLSVKRRCEPVLKEFGFAWPDSLNCSEKFPQNDHNHMCMEGPGDEEVLPKHTPIQPG
ECHSVGTNSDQYIWKRLNLCVLCGYDAGLYSRSAKEFTDIWMAVWASLCFISTTFTVL
TFLIDSSRFSPYERPIIFLSMCYNIYSIAYIVRLTVGRERISCFEAAEPVLIQGLKN
20 TGCAIIFLLMYFFGMASIIWVILTLTWFLAAGLKWGHEAIEMHSSYFHIAAWAIPAVKT
IVILIMRLVDADELTLGCLYVGNQSLDALTGFFVAPLFTYLVIGTLFIAAGLVALFKIRSN
LQKDGTKTDKLERLMVKIGVFSVLVTPATCVIACYFYEISNWALFRYSADDSNMAVEML
KIFMSLLVGITSGMWISAKTLHTWQKCSNRLVNSGKVKREKRGNGVWKPGKGNETVV

```

SEQ ID NO: 117

25 gi|5803150|ref|NM\_006851.1| Homo sapiens GLI pathogenesis-related 1 (glioma) (GLIPR1), mRNA

```

1 ctctgttttc tcaaagctga agtcggctag gtttgcaaag ctgtgggctg agcactcagg
61 caatcacact ctcagaaact gcggcggtc tggactgcag cctcccaagg ctccatgcc
121 gacaaagcat gcgtgtcaca cttgtctaca tagcctggat ggtttctttt gtctccaatt
30 181 attcacacac agcaaatatt ttgccagata tcgaaaatga agatttcacg aaagactcgg
241 ttcgaatcca taacaagttc cgatcagagg tgaaaccaac agccagtgat atgctataca
301 tgacttggga cccagcacta gcccaaattg caaaagcatg ggccagcaat tgccagtttt
361 cacataatac acggctgaag ccaccccaca agctgcaccc aaacttcact tcaactggag
421 agaacatctg gactgggtct gtgcccattt tttctgtgtc ttccgccatc acaaaactggt
35 481 atgacgaaat ccaggactat gacttcaaga ctcgatatg caaaaaagtc tgtggccact
541 acactcagggt tggttgggca gatagttaca aagttggctg cgcagttcaa ttttgccta
601 aagtttctgg ctttgacgct ctttccaatg gagcacattt tatatgcaac tacggaccag
661 gaggggaatta cccaacttgg ccatataaga gaggagccac ctgcagtgcc tgccccaata
721 atgacaagtg tttggacaat ctctgtgtta accgacagcg agaccaagtg aaacgttact
40 781 actctgttgt atatccaggc tggcccatat atccacgtaa cagatacact tctctctttc
841 tcattgttaa ttcagtaatt ctaatactgt ctgttataat taccattttg gtacagctca
901 agtaccctaa tttagtctt ttggactaat acaattcagg aaagaaaaaa cccaaaaacc
961 aacctcattc acatatggct ttttttaac caataacaat taggtgtact tctattttta
1021 aacatttcag aaaaaaatat atgttatagc aatactctta c

```

45 SEQ ID NO: 118

Amino acid sequence of human GLIPR1 encoded by the DNA sequence shown in SEQ ID NO: 117.

MRVTLATIAWMVSFVSNYSHTANILPDIENEDFIKDCVRIHNKFRSEVKPTASDMLYMTW

DPALAQIAKAWASNCQFSHNTRLKPPHKLHPNFTSLGENIWTGSPVIFSVSSAITNWDYDE  
IQDYDFKTRICKKVCGHYTVVWADSYKVGCAVQFCPKVSGFDALSNGAHFICNYGPGGN  
YPTWPKRGATCSACPNNDKCLDNLCVNRQRDQVKRYYSVVYPGWPIYPRNRYTSLFLIV  
NSVILILSVIITILVQLKYPNLVLLD

## 5 SEQ ID NO: 119

gi|21312071|ref|NM\_028608.1| Mus musculus GLI pathogenesis-related 1 (glioma) (Glipr1), mRNA

```

1 gagcatgctg aagatggagc tcagaggcag agcacttgct tagcataaac aacctggggt
61 taatccgagc tccaacaggg aaacagtctg cagactgaga gaaccgagca ttctatcaga
10 121 accccgcagc tctggattct aggtccagca gcaaccagag agaccatgca gggtcatcctt
181 gctgtgatag tctggatggc ttcgtctgtg tctagtctct catttacagc aagcactttg
241 ccagatataa caaacgagga cttcattaaa gaatgtgttc aagttcaca ccagcttcgg
301 tcaaaagtga gtccaccagc ccggaatatg ctgtacatgt cttgggaccc aaaactagcc
361 caaattgcaa aagcatggac aaaatcttgt gaatttaaac acaaccaca gctgcattca
15 421 cggatacacc caaatttcac cgccctggga gagaatatct ggcttggctc tctatccatc
481 ttttcagtat cctcagccat ctctgcctgg tatgaagaaa ttaagcacta tgacttcagc
541 actaggaaat gtagacatgt ctgtggccat tatactcagg ttgtttgggc agacagttac
601 aaacttggct gtgcagtga actttgccct aatggagcaa attttatatg cgactatgga
661 ccagcaggaa attacccaac gtggccatat aagcaaggag ccacgtgcag tgattgccc
20 721 aaagatgaca agtgtctcaa cagtctctgc attaacccac gacgagacca ggtctcacgt
781 tactactctg tcgattatcc agactggcct atatacctgc gtaacagata cacatctctc
841 tttctcattg ctaagtcggt tctcctatta ctgtctgtta taattaccat ctgggtaaa
901 cacaatatc ctaacttggg tcttttgga taaagctgtg gttgggggac aactgaatca
961 catgcggcta tttaaaaact tttcaataaa atctcagtca aaagg

```

## 25 SEQ ID NO: 120

Amino acid sequence of mouse GLIPR1 encoded by the DNA sequence shown in SEQ ID NO: 119.

```

30 MQVILAVIWMASVSSSSFTASTLPDITNEDFIKECVQVHNQLRSKVSPARNMLYMSW
DPKLAQIAKAWTKSCEFKNPQLHSRIHPNFTALGENIWLGSLSIFSVSSAISAWYEEIK
HYDFSTRKCRHVCGHYTVVWADSYKLGCAVQLCPNGANFICDYGPAGNYPTWPKQGAT
CSDCPKDDKCLNSLCINPRRDQVSRYYSVDYPDWPIYLRNRYTSLFLIAKSVLLLLSVII
TIWVKHKYPNLVLLD

```

## SEQ ID NO: 121

gi|27718248|ref|XM\_216892.1| Rattus norvegicus similar to GLI pathogenesis-related 1 (glioma); related to testes-specific, vespid, and pathogenesis proteins (LOC299783), mRNA

```

1 gtgaactgaa gatggggctc aggagcagag cacttgctta gcataaaca ccttgggttc
61 attcaagctc caacatggaa actgtctgca gactgagcga accaagcatt ctatcagaac
121 cctgaggccc tggattctag gtccagcagc agccagaggg accatgcagg tctctctgc
181 tgtgatggtc tggatggctt cttctgcgtc tgggttttca tatacagcaa gtactttgcc
40 241 aaaaataaca aacgaggact tcacgaaga atgcgttgaa gttcacaacc actttcggtc
301 aaaagcgtat ccaccggccg ggaatatgtt gtacatgtct tgggacccaa aactagccca
361 aattgcaaaa gcgtgggcac agtcttctgt atttcaacac aaccacagc tgcattcgcg
421 aatacaccca aactttactg gcctgggcga aaacatttgg cttggctctc tatccctctt
481 ctacgtacgt gcggccatcc tcgcttggtt tgaagaaagc cagtactatg acttcagcac
45 541 tgggaaatgt aaaaaagtct gtgggcatta cactcagatt gtttgggcag atagttacaa
601 gattggctgt gcagtgaac tctgcccagc aggagcaaat tttatttgca actatggacc
661 agcaggggat taccacacgt ggccatataa gcagggagcc acttgcagtg cttgcccaca
721 ggatgacaag tgcctgaaca atctctgcac taaccacaaa cgagatcagg tctcacgtca

```

781 ctctgctgat tatccaaaat atctacgtaa cagatacaca tcgctctatc tcatcgctaa  
 841 gtcagttctc ctattactgt ctgtcataat taccatttgg gtaaagcaca aatatacctaa  
 901 cttagtctct ttggactaat gccctgggta ggggacaact tattcatatg tgggtgatttt  
 961 aaaatgtttc aataaaaatct tagaagagtt

## 5 SEQ ID NO: 122

Amino acid sequence of rat GLIPR1 encoded by the DNA sequence shown in SEQ ID NO: 121.

10 MQVLLAVMVWMASSASGFSYASTLPKITNEDFIEECVEVHNHFRSKAYPPAGNMLYMSW  
 DPKLAQIAKAWAQSCVFQHNPLHSRIHPNFTGLGENIWLGLSLFVRAAILAWFEESQ  
 YYDFSTGKCKKVCGHYTOIVWADSYKIGCAVQLCPRGANFICNYGPAGNYPTWPYKQGAT  
 CSACPKDDKCLNNLCTNPQRDQVSRHSADYPKYLNRNRYTSLYLIKSVLLLLSVIITIWW  
 KHKYPNLVLLD

## SEQ ID NO: 123

15 gi|7661847|ref|NM\_014879.1| Homo sapiens G protein-coupled receptor 105 (GPR105), mRNA

1 gaacagtgtt accttggagc ctacaatgag aggtatttca aaatgagtga agcatgactc  
 61 tcacagatga aggcctagac gcaggatctt taatggaaaa acacttgggc cacttcaaga  
 121 cgacaaacgc tcaactgggca aaacaccttc actgaaaaga gacctcatat tatgcaaaaa  
 181 aaatcttaag aggcctctgc cttcagaagt tacaagatga tcaattcaac ctccacacag  
 20 241 cctccagatg aatcctgtctc tcagaacctc ctgatcactc agcagatcat tcctgtgctg  
 301 tactgtatgg tcttcattgc gggaatccta ctcaatggag tgtcaggatg gatattcttt  
 361 tacgtgcccc gctctaagag tttcatcatc tatctcaaga acattgttat tgcgtacttt  
 421 gtgatgagcc tgacttttcc tttcaagatc cttggtgact caggccttgg tccctggcag  
 481 ctgaacgtgt ttgtgtgcag ggtctctgcc gtgctcttct acgtcaacat gtacgtcagc  
 25 541 attgtgttct ttgggctcat cagctttgac aggtattata aaattgtaaa gcctctttgg  
 601 acttctttca tccagtcagt gaggttacagc aaacttctgt cagtgatagt atggatgctc  
 661 atgctcctcc ttgctgttcc aaatattatt ctacccaacc agagtgttag ggaggttaca  
 721 caaataaaat gtatagaact gaaaagtga ctgggacgga agtggcaca agcatcaaac  
 781 tacatcttcg tggccatctt ctggattgtg tttcttttgt taatcgtttt ctatactgct  
 30 841 atcacaaaga aaatctttta gtccacactt aagtcagatc ggaattccac ttccgtcaaa  
 901 aagaaatcta gccgcaacat attcagcatc gtgtttgtgt tttttgtctg ttttgtacct  
 961 taccatattg ccagaatccc ctacacaaag agtcagaccg aagctcatta cagctgccag  
 1021 tcaaaagaaa tcttgcggtg tatgaaagaa ttcactctgc tactatctgc tgcaaatgta  
 1081 tgcttggacc ctattattta tttctttcta tgccagccgt ttagggaat cttatgtaag  
 35 1141 aaattgcaca ttccattaaa agctcagaat gacctagaca tttccagaat caaaagagga  
 1201 aatacaacac ttgaaagcac agatactttg tgagttccta ccctcttcca aagaaagacc  
 1261 acgtgtgcat gttgtcatct tcaattacat aacagaaatc aataagatat gtgccctcat  
 1321 cataaatatc atctctagca ctgccatcca atttagttca ataaaattca aatataagtt  
 1381 tccatgcttt tttgtaacat caaagaaaac atacccatca gtaatttctc taatactgac  
 40 1441 ctttctattc tctattaata aaaaattaat acatacaatt attcaattct attatattaa  
 1501 aataagttta agttttataac cactagtctg gtcagttaat gttagaaattt aaatagtaaa  
 1561 taaaacacaa cataatcaaa gacaactcac tcaggcatct tctttctcta aataccagaa  
 1621 tctagtatgt aattgttttc aacactgtcc ttaaagacta acttgaaagc aggcacagtt  
 1681 tgatgaaggg cttagagact gtttgcaata aaaagtcagg ttttttctct gatttgaaga  
 45 1741 agcaggaaaa gctgacaccc agacaatcac ttaagaaacc cttattgat gattttcatg  
 1801 gcactgcaaa ggaagaggaa tattaattgt atacttagca agaaaatttt ttttttctga  
 1861 tagcactttg aggatattag atacatgcta aatatgtttt ctacaaagac ttacgtcatt  
 1921 taatgagcct ggggttcttg tgttagaata tttttaagta ggctttactg agagaaacta  
 1981 aatattggca tacgttatca gcaacttccc ctgttcaata gtatgggaaa aataagatga  
 50 2041 ctgggaaaaa gacacacca caccgtagaa catatattaa tctactggcg aatgggaaag  
 2101 gagacatttt tcttagaaag caataaaact tgattttttt aaatctaaaa ttacatttaa  
 2161 tgagtgcata ataacacata aaatgaaaat tcacacatca catttttctg gaaaacagac



2221 ggatttttact tctggagaca tggcatacgg ttactgactt atgagctacc aaaactaaat  
 2281 tcttttctctg ctattaactg gctagaagac attcatctat ttttcaaag ttctttcaaa  
 2341 acatttttat aagtaatgtt tgtatctatt tcatgcttta ctgtctatat actaataaag  
 2401 aaatgtttta atactg

## 5 SEQ ID NO: 124

Amino acid sequence of human GPR105 encoded by the DNA sequence shown in SEQ ID NO: 123.

10 MINSTSTQPPDESCQNLLITQQIIPVLYCMVFIAGILLNGVSGWIFFYVPSSKSFIIYL  
 KNIVIADFVMSLTFFPKILGDSGLGPWQLNVFVCRVSAVLFYVNMYSIVFFGLISFDY  
 YKIVKPLWTSFIQSVSYSKLLSVIVWMLMLLLAVPNIIILTNQSVREVTVQIKCIELKSELG  
 RKWHKASNYIFVAIFWIVFLLLIVFYTAITKKIFKSHLKSRRNSTSVKKKSSRNIFSIVF  
 VFFVCFVPYHIARIPYTKSQTEAHYSCQSKEILRYMKEFTLLLSAANVCLDPIIYFFLCQ  
 PFREILCKKLHIPLKAQNLDLISRIKRGNTTLESTDTL

## SEQ ID NO: 125

## 15 gi|34328344|ref|NM\_133200.2| Mus musculus G protein-coupled receptor 105 (Gpr105), mRNA

1 aattcggatc catggaaggc cgccacccca gcagactgaa gccagacgtg aaggagttca  
 61 tgtaagggag tccctgctgt cctccagaca cactgatgcc tgggctacgg atggggacgg  
 121 ggacgcaatg tgtctggaat tctctcttcc gaatcctgga ttctgttgac gaagcttgcc  
 20 181 tttgagattc ctgaacacgg agaaatagag attaaaaacc ccagaagaga gaaagtaa  
 241 gattcacaaat cttgatgggt tttgccgtat ttatgttctt cactgtatt agataccagt  
 301 cacaaatgac ttagaggcca taaactgtgc tttaagtaac tagcctgcct ttctatccag  
 361 atctttgcct ccagagggtga gaagatgaac aactccacca ccacagaccc tccaaaccag  
 421 ccctgtctct ggaacaccct gatcacaag cagatcattc cgtgtgtgta cggtaggtc  
 25 481 ttcacacagg ggctcctcct caatgggata tcaggatgga tattctttta tgtgccagc  
 541 tccaagagtt ttatcatcta tctcaagaac atagtgggtg ctgactttct catgggcctg  
 601 actttccctt tcaaagtctt tggtagtca ggcctcggcc cctggcaggt gaatgtgtt  
 661 gtgtgcaggg tctctgccgt catcttctat gttaatatgt acgtcagcat cgtgttctt  
 721 gggctcatga gctttgacag gtactataaa attgtgaagc cccttctgac gtctattgtg  
 30 781 cagtgcgtga actatagcaa gctgctttct gtgctcgtgt ggatgtcat gcttctcct  
 841 cgtgtcccaa acatcatcct gacaaaccag ggtgtcaagg aggtcacgaa gatacagtgc  
 901 atggagctca aaaacgagct ggggcggaag tggcacaagg cgtctaacta tatcttcgtg  
 961 agtatcttct gggctcgtgtt tcttctgcta atcgtcttct acacggccat cacgaggaag  
 1021 atcttcaagt ctcacctcaa gtccaggaag aattccacct ccgtcaagag gaagtccagc  
 35 1081 cgcaatatct tcagcatcgt gctcgttttt gtcgtctgct ttgtgcctta ccacattgcc  
 1141 agaatcccc acacaaagag tcagacggaa ggtcactaca gctgccggac gaaggagacc  
 1201 ctgctctatg cgaaagaatt cactctgcta ctctcggctg ccaatgtgtg tctggacccc  
 1261 attatttatt tcttcttatg ccagccattt agagaagtct taaataagaa gttacacatg  
 1321 tcaactcaaag tccagaatga cctagagggt tccaaaacca aaagggaata tgcgattcat  
 40 1381 gaaagcacag atactttgta aattcccatc cccttccaag tattatcagt cttgttacat  
 1441 gataattaag atacatgaat aaaaagcagg catatgatga taagtaactt agctagcaat  
 1501 atatctaata atatgtatga agtccaaaaa ggtataataa aaataaaaata taagtttcca  
 1561 tgcaaaatgg aagtntgtag cacatcacat ttttttagaa atcaaaaggaa cagagaagtg  
 1621 gctttgtggg tgctggcgta tgagttacca aaaccaaact tctcttctat taactggett  
 45 1681 cttagaagac acccagctct tccgaccttc ctctaagca ttcttccaag caacactcgt  
 1741 atctatttca tgctttgtac tatgcattgtg ccaataaaca agttgtcttc aaacccaaa  
 1801 aaaaaaaaaa aaaaaaaaaa agggcgcccg caagcttatg tatttaa

## SEQ ID NO: 126

Amino acid sequence of mouse GPR105 encoded by the DNA sequence shown in SEQ ID NO: 125.

5 MNSTTTDPPNPQCSWNTLITKQIIPVLYGMVFITGLLNGISGWIFFYVPSSKSFIIYL  
KNIVVADFLMGLTFPFKVLGDSGLGPWQVNVFVCRVSAVIFYVNMVVSIVFFGLISFDRI  
YKIVKPLLTISIVQSVNYSKLLSVLVWMLMLLLAVPNIIILTNQGVKEVTKIQCMEKKNELG  
RKWHKASNYIFVSIFWVVFLLLIIFYTAITRKIFKSHLKSRRNSTSVKRKSSRNIFSIVL  
VFVVCVFPYHIARIPYTKSQTEGHYSCRTKETLLYAKEFTLLLSAANVCLDPIIYFFLCQ  
PFREVLNKKLHMSLVQNDLEVSCTKRENAIHESDTL

SEQ ID NO: 127

10 gi|25742688|ref|NM\_133577.1| Rattus norvegicus G protein-coupled receptor 105 (Gpr105), mRNA

1 aattctctct tccgcaccct ggggtctgtg gatgaaactt gcctctgaga atcctgaaca  
61 tggagaaata gagataaata cccagagaaga gggggagtaa gtgttcacaa ccttggtggg  
121 tttcgctcca tgtctgtccc tccactttaa gagatgccgg tcacctgagg gccacaaact  
15 181 gcgctctaag taaccagcct gcctttctac ccagatcttt gtctccagaa gtgagaagat  
241 ggacaacaca acaaccacag aacctccaaa gcagccctgc acccggaaca cctgatcac  
301 acagcagatc atccccatgt tgtactgtgt ggtcttcac acaggggtcc tctcaacgg  
361 aatctcgga tggatattct tttacgtgcc cagctctaag agttttatca tctatctcaa  
421 gaacatagtt gtggctgact ttctcatggg cctcactttc ctttcaaaag tctcagcga  
20 481 ctacggcctc ggtccctggc agctgaatgt gttgtattc aggggtgtctg ccgtgatctt  
541 ctacgtcaac atgtacgtca gcacgcgtt cttcgggctc atcagctttg acaggtacta  
601 taaaatcggtg aagcccttc tgggtgtctat cgtccagtca gtgaactaca gcaaagtgtc  
661 gtccgtgctt gtgtgggtgc tcatgtctct cctcgtgtgc cccaacatca ttctgacaaa  
721 ccagagtgtc aaggatgtca ctaacataga gtgcatggaa ctcaaaaacg agctggggcg  
25 781 gaagtggcac aaggcgtcta actatgtctt cgtgagcatt ttctggatcg tgttccttct  
841 gctgaccgtc ttctacatgg ccataacgag gaagatcttc aagtctcacc tcaagtccag  
901 aaagaattcc atctccgtca aaaggaagtc cagccgcaat atattcagca ttgtgtctgc  
961 atttgtcgcc tgtttcgcc cttaccagtc tgccaggatc cctacacca agagccagac  
1021 ggaaggacac tacagtgccc aggcacaagg gacctgctc tatacgaag aattcacct  
30 1081 gctgctctcg gctgccaatg tgtgtctgga ccccatatct atttcttct atgccagccg  
1141 tttagagaag tcttgaataa gaagtaaga atgtactaa cagtcacaga tgacctagag  
1201 acttccaaaa ccagaagggg aaatatgatt caggaaagca cagatacttt gcaattctca  
1261 ccctttccaa gtattatgga cctgtgtaca cggtaattaa gatgtatgaa gtgaaaagca  
1321 gaaaagtata ataaaaatga aaggttaagt cccatgtaaa gtgaaagtct acagcacgca  
35 1381 gaaaagtata ataaaaatga aaggttaagt cccatgtaaa gtgaaagtct acagcacgct  
1441 acatcttttt agaaatcgaa tggagaagtg gcatgtgggt gctagtctgc gagttacccc  
1501 aaaactaaac tctctcttct attaacgggc ttcttagaag acaccacccc ttccagaagt  
1561 tctctctgag tctcttcta agtaatgctc acctctgtt catgctttgt tctgtgtatg  
1621 cgccaataaa caagtgtgtc ttaaaacccc aaa

40 SEQ ID NO: 128

Amino acid sequence of rat GPR105 encoded by the DNA sequence shown in SEQ ID NO: 127.

45 MDNTTTTEPPKQPCTRNLTITQIIPMLYCVVFITGVLLNGISGWIFFYVPSSKSFIIYL  
KNIVVADFLMGLTFPFKVLSDSGLGPWQLNVFVFRVSAVIFYVNMVVSIAFFGLISFDRI  
YKIVKPLLVSIVQSVNYSKLVSVLVWVLMMLLLAVPNIIILTNQSVKDVNTIQCMEKKNELG  
RKWHKASNYVFSIFWIVFLLLTVFYMAITRKIFKSHLKSRRNSISVKKSSRNIFSIVL  
AFVACFAPYHVARIPYTKSQTEGHYSCQAKETLLYTKFTLLLSAANVCLDPISSISYAS  
RLEKS

SEQ ID NO: 129

gi|19923974|ref|NM\_138445.1| Homo sapiens G protein-coupled receptor 146 (GPR146), mRNA

```

      1  ggcacgaggc gccggcgcgc atgtggagct gcagctgggt caacggcaca gggctgggtgg
    5  61  aggagctgcc tgcctgccag gacctgcagc tggggctgtc actgttgctg ctgctggggcc
      121 tgggtgggtggg .cgtgccagtg ggctgtgtct acaacgccct gctgggtgctg gccaacctac
      181 acagcaaggc cagcatgacc atgccggacg tgtactttgt caacatggca gtggcaggcc
      241 tgggtgctcag cgccctggcc cctgtgcacc tgcctggccc cccgagctcc cggtagggcgc
      301 tgtggagtggt gggcgggcgaa gtccacgtgg cactgcagat ccccttcaat gtgtcctcac
      361 tgggtggccat gtactccacc gccctgtgta gcctcgacca ctacatcgag cgtgcactgc
   10  421 cgcggaccta catggccagc gtgtacaaca cgcggcacgt gtgcggtctc gtgtggggtg
      481 gcgcgctgct gaccagcttc tcttcgtctc tcttctacat ctgcagccat gtgtccaccc
      541 gcgcgctaga gtgcgccaag atgcagaacg cagaagctgc cgacgccacg ctggtgttca
      601 tcggctacgt ggtgccagca ctggccaccc tctacgcgct ggtgctactc tcccgcgtcc
      661 gcagggagga caccgcccctg gaccgggaca cggggcgggt ggagccctcg gcacacaggc
   15  721 tgctgggtggc caccgtgtgc acgcagtttg ggctctggac gccacactat ctgatcctgc
      781 tggggcacac ggtcatcatc tcgcgaggga agcccgtgga cgcacactac ctggggctac
      841 tgcactttgt gaaggatttc tccaaactcc tggccttctc cagcagcttt gtgacaccac
      901 ttctctaccg ctacatgaac cagagcttcc ccagcaagct ccaacggctg atgaaaaagc
      961 tgccctgcgg ggaccggcac tgctccccgg accacatggg ggtgcagcag gtgctggcgt
  20 1021 aggcggccca gccctcctgg ggagacgtga ctctggtgga cgcagagcac ttagttacce
      1081 tggacgctcc ccacatcctt ccagaaggag acgagctgct ggaagagaag caggaggggt
      1141 gtttttcttg aagtttctct tttcccacaa atgccactct tgggccaagg ctgtgggtccc
      1201 cgtggctggc atctggcttg agtctccccg aggcctgtgc gtctcccaaa cacgcagctc
      1261 aaggtccaca tctgcaaaag cctcctcgcc ttcagcctcc tcagcattca gtttgtcaat
  25 1321 gaagtgatga aagcttagag ccagatttta tactttgtgg ttaaaatact tgattcccc
      1381 ttgtttgttt taaaaaaca gatgtttcct agaaaaatga caaatagtaa aatgaacaaa
      1441 accctacgaa agaatggcaa cagccagggt ggccggggcc tgccagtggg cggcgtgtgc
      1501 tagcaaggcc tgccgggtgt gccgcagtca ccacagggtt ctgagaacat ttcacagaag
      1561 tgctgagac gcggagacat ggctgggtgt aaatggagct attcaatagc agtgacgcgc
  30 1621 tctcctcagc caccaaatgt ccctgacacc ctccccagcc cccacagata acatcagctg
      1681 aggttttttt cagtatgaac ctgtctaaa tcaattcttc aaagtgtgca caaaactaaa
      1741 gaatataaat aaacaaaaga aaggtgaaaa aaaaaaaaaa aaaaa

```

SEQ ID NO: 130

Amino acid sequence of human GPR146 encoded by the DNA sequence shown in SEQ ID NO: 129.

```

  35  MWSCSWFNGTGLVEELPACQDLQLGLSLLSLLGLVVGVPVGLCYNALLVLANLHASKASMT
      MPDVYFVNMAVAGLVLSALAPVHLLGPPSSRWALWSVGGEVHVALQIPFNVSSLVAMYST
      ALLSLDHYIERALPRTYMASVYNTRHVC GFVWGGALLTSFSSLLFYICSHVSTRALECAK
  40  MQNAEADATLVF IGYYVPALATLYALVLLSRVRREDTPLDRDTGRLEPSAHRLLVATVC
      TQFGLWTPHYLILLGHTV IISRGKPVDAHYLGLLHFVKDFSKLLAFSSSFVTPLLYRYMN
      QSFPSKLQRLMKKLPCGDRHCS PDHMGVQQLA

```

SEQ ID NO: 131

gi|31981401|ref|NM\_030258.2| Mus musculus cDNA sequence BC003323 (BC003323), mRNA

```

  45  1  ggtacacagc cccacggcg tgccgcgcgt gagctccgct gcctcctgca agcccagctg
      61  ccgcccgcgc gccggtgcgc gggccgcccgc ccagcgagcg ctgcgaagct ggggtgagct
      121 gccatgtgga gctgtggccc actcaacagc acagcgtggg ctgaggagcc gctgtgccgg
      181 aacctgcgcc tggggctgtg ggtcctctcg ctgctctacc tgggggcagg ggtccctgtg
      241 agcttaggct ataatgctct tctggtgctg gccaacctgg ccagcaagaa caccatgacc
  50  301 atgcccagc tgacttcgt gaacatggcc gtggcggggc tgggtgctac ggcactggca
      361 cctgcgtacc tggtggggcc tgcccactcc aggtggggcc tgtggagcct cagcagtga

```

```

421 gccatgtga cactgctcat cctgttcaac gtggcttccc tggtagccat gtactccact
481 gcaactgctga gccttgacta ctacatcgag cgtgcccctgc cacgcacctta catggccaggt
541 gtgtacaaca cccggcacgt gtgtggcttc gtctggggag gggcggtgct caccagcttc
601 tectccctgc tcttctacat ctgcagtcac gtgtcttcta gaatcgctga gtgtgcccgg
5 661 atgcagaaca cggaggcagc cgatgctatc cttgtgctca tcggctacgt ggtgccaggt
721 ctggctgtgt tgtatgccct ggcaactcatc tcgagaatcg ggaaggaaga cacacccctg
781 gaccaggaca ccagcaggct ggacccctcg gtgcacaggc tgcgtggtggc caccgtgtgc
841 actcagtttg gcctctggac accttactac ttgagcctgg ggcacacagt gctgacgtca
901 cgggggagga ccgtggaggg gcattatctg ggcacacctac aggttgctaa ggacctggcc
10 961 aagttcctgg ccttctcaag cagttctgtg acaccactgc tctaccgtta catcaacaaa
1021 gccttccccg gcaagctccg gcggctgatg aagaagatgc actgcgggag ccgccactgc
1081 tcccccgacc cctcggggat acagcagggt atggcacagg cgtagctaac cctccctggg
1141 ctgaccacga ggaacactaa aactcaactg gacacatagc actttgttcc ccaggcacat
1201 ggagcttcca gccaggaga cagcgtgaag ccagagatgc acagcggggc atttctcttg
15 1261 atgtttctg ttttccccta aaaggccaac cttggtctta tgcgtcatgg tggaaagcag
1321 aggtgccctg tgtgaacagg ctgcttctga ctaaccttcc cacttctctc ctcaagggtg
1381 gtgcctttct cccagggaca tcttcaaggt ccatgtcttc acaagcaggc tgaccttcgg
1441 cctccctggg atttgggctt tctaaatgaa gcgatgaaat ctaaagccag tatttatact
1501 tcatattgac atgatacttg attctcctat tattattttt ttaaaaatag aagtgttaga
20 1561 aagataccta gtattaagaa gaaggcactt caggggaatg tggctgtcgc tgataccagt
1621 gttagagctt gaggacagga cagtgggctt gtcaggcctg cagcagctac acagccattg
1681 ctgacagtga cagctatgga gcactggccc tctcctctac tgagagagga tctttccagt
1741 gcttcaacat atgcatat

```

SEQ ID NO: 132

- 25 Amino acid sequence of mouse GPR146 encoded by the DNA sequence shown in SEQ ID NO: 131.

```

MWSCGPLNSTAWAEPLCRNLRLGLWVLSLLYLGAGVPVSLGYNALLVLANLASKNTMTM
PDVYFVNMAVAGLVLTALAPAYLLGPAHSRWALWSLSSEAHVTLILLFNVASLVMTMYSTA
LLSLDYIERALPRTYMASVYNTRHVCGFVWGGAULTSFSSLLFYICSHVSSRIACARM
30 QNTEAADAILVLIGYVVPGLAVLYALALISRIGKEDTPLDQDTSRLDPSVHRLLVATVCT
QFGLWTPYYLSLGHVLTSLRGRTVEGHYLGILQVAKDLAKFLAFSSSSSVTPLLYRYINKA
FPGKLRRLMKMKHCGRRHCSPPDSGIQQVMAQA

```

SEQ ID NO: 133

ENSRNOT00000001733 cDNA sequence, EnsEMBL transcript [Rattus norvegicus]

```

35 1 atgtggagct gtggccact caacagcaca gcgtgggctg aggagccgct gtgccggaac
61 ctgcgcctgg ggctgtgggt cctctcactg ttctacctgg gggcagggtg ccctgtgggc
121 ttaggctaca atgtctcttt ggtgctggcc aatctggcca gcaagaacag catgaccatg
181 cctgatgtgt acttcgtgaa catggctgtg gcggggctgg tgctcacagc actggcacct
40 241 gcgtacctgc tgggtcctgc ccactccagg tgggccctgt ggagcctcag cagcagggcc
301 catgtgacac tgctcatcct gttcaacgtg gcttccctgg tgacctatga ctccactgca
361 ctgctgagtc ttgactacta categagcgt gccctgccgc gtacctacat ggctagtgtg
421 tacaacaccc ggcacgtgtg tggcttcgtc tggggagggg cagtgtctac cagcttttcc
481 tccctgctct tctatatctg cagtcatgtg tcttctagaa ttgcccagtg tgcccggatg
541 cagaacacgg aggcagccga cggcatcctt gtgctcattg gctacgtggg gccaggctctg
45 601 gctgtgttgt atgccctggc actcatctca aggattggga aggaagacac acccctggac
661 caggacacca gcaggctgga cccctcagtg cacaggctgc tggtagccac tgtgtgcaca
721 cagtttggcc tctggacacc ttactacctg agcctggggc acacagtgtt agtgtcacgg
781 ggaaggaccg tggaggggca ttatctgggc atctacagc ttgctaagga cctggcaaaag
841 ttcttggcct tctcaagcag cctgtgtgac cgtgtgtctt accgttacat caacaaagcc
50 901 tccccagca agctccggcg cctgggtgaaa aagatacact gtgggcggcg ccactgctcc
961 cccgaccctg cggggatata gcagggtgat gccaggcgt ag

```

SEQ ID NO: 134

Amino acid sequence of rat GPR146 encoded by the DNA sequence shown in SEQ ID NO: 133.

5 MWSCGFLNSTAWAEEPLCRNLRLGLWVLSLFYLGAGVPVGLGYNALLVLANLASKNSMTM  
PDVYFVNMAVAGLVLTALAPAYLLGPAHSRWALWSLSSEAHVTLILFNVASLVTMYSTA  
LLSLDYIERALPRTYMASVYNTRHVCGFVWGAVLTSFSSLLFYICSHVSSRIAECARM  
QNTAADAILVLIGYVVPGLAVLYALALISRIKEDTPLDQDTSRLDPSVHRLLVATVCT  
QFGLWTPYYLSLGHYTLVSRGRTVEGHYLGILQVAKDLAKFLAFSSSSVTPLLYRYINKA  
FPSKLRLRLVKKIHCGRRHCSPPDAGIQVMAQA

10 SEQ ID NO: 135

gi|4504090|ref|NM\_001505.1| Homo sapiens G protein-coupled receptor 30 (GPR30), mRNA

1 ggaaaacgac acctagaagt aggagtgaga ttcgctgaag ttcccttctg aggaagaccc  
61 acccctccgc ctggagagcc ggggctggcg gtgcctgagg accccttcgg cctggacagc  
121 ccacgcgggc ttggggggcc tcgctctgcc ctcatggggc ggccatcggt tccgaagcg  
15 181 gcgagtgaat attcaaatgg ccagtagggg gcgcactcgg aagtggccgc cccgcagtag  
241 gcagttcagc ggccccgaga gtccggggag ggaggtttat tctccgcctg cagcagactg  
301 tgaaatccgc aaccatgagc aggagaggcg gccctggtgg ggaagaggcc accaaccatct  
361 ggacggcagg taccagaga gtgagcagct ccacgcggga ctgtgcacgg tggccgacac  
421 ccgcaggagc gcccgcgga cagacacgcg gagggccctc gcctccacgg atgcaccatg  
20 481 ccggtgtgag gacatctgt tcttccact ctctgcagt aacaaacca acccaacca  
541 ccacagggtg tcttctctgg gaggtttctg tctgacaaat gccaggctca cttcaaggag  
601 aatcacgctt ctttctaaag atggattcac catttaaac agagctctgg gaggcttctg  
661 gcaaatcttg aaagctgcac ggcgcagaga catggatgtg acttcccaag cccggggcgt  
721 gggcctggag atgtaccag gcaccgcgca gcctgcggcc cccaacacca cctccccga  
25 781 gctcaacctg tcccaccgc tcttgggac cgcctgggac aatgggacag gtgagctctc  
841 ggagcaccag cagtacgtga tcggcctgtt cctctcgtgc ctctacacca tcttctctt  
901 cccatcggc tttgtgggca acatctgat cctggtggtg aacatcagct tccgcgagaa  
961 gatgaccatc cccgacctgt acttcatcaa cctggcggtg gcggacctca tctggtggc  
1021 cgactccctc attgaggtgt tcaacctgca cgagcggtag tacgacatcg ccgtcctgtg  
30 1081 caccttcatg tcgctcttcc tgcaggtaa catgtacagc agcgtcttct tctcaccctg  
1141 gatgagcttc gaccgctaca tcgccctggc caggggcatg cgctgcagcc tgttccgcac  
1201 caagcaccac gcccggctga gctgtggcct catctggatg gcacccgtgt cagccacgct  
1261 ggtgcccttc accgcccgtg acctgcagca caccgacgag gcctgcttct gtttcgaggc  
1321 tgtccgggag gtgcagtggc tcgaggtcac gctgggcttc atcgtgccct tcgccatcat  
35 1381 cggcctgtgc tactccctca ttgtccgggt gctggtcagg gcgcaccggc accgtgggct  
1441 gcggcccccg cggcagaagg cgtccgcag gatcctcgcg gtggtgctgg tcttctctgt  
1501 ctgctggctg ccggagaacg tcttcatcag cgtgcacctc ctgcagcgga cgcagcctgg  
1561 ggccgctccc tgcaagcagt ctttccgcca tgcccacccc ctcacgggac acattgtcaa  
1621 cctcgccgccc ttctccaaca gctgcctaaa cccctcctac tacagcttct tccgggagac  
40 1681 cttcaggggc aagctgaggc tgtacattga gcagaaaaca aatttgccgg ccctgaaccg  
1741 cttctgtcac gctgccctga aggcgctcat tccagacagc accgagcagt cggatgtgag  
1801 gttcagcagt gccgtgtaga cagccttggc cgcataggcc cagccagggt gtgactcggg  
1861 agctgcacac acctgggtgg acacaaggca cggccacgct atgtctctaa actgagggtc  
45 1921 gatgtggctt ctggctcctc ggggcctcgc gagggtcacg cttgcctggt caccctgggg  
1981 ctgcttagga aacctcacga ctggtcacct tgcactcttc acacagaatt gctacaatcc  
2041 caaagcgctc gcccgcgagg gtccaaaggc cagcgggtgac cagccctgca cccagctcct  
2101 ccccgcacac cctgcctgcc gctgcacctg cctgccgctg caggaaacat ttgacaccgt  
2161 cgaccaggaa agccacacgg agaggccact gtgggtgaag cgctcagtt acacaggaa  
2221 cctaaagcaa atctgccacc tggggggaac tgacgctgga gatgcaagg gtggtgggtg  
50 2281 ctgagctgga cgtcgcggtg tgtcctctgt gccacgggc tgagctagct agcgcaccgc  
2341 cgagttaaag aggagaagga aaacatgctg ctctggtgca cgctgagcg tctccatct  
2401 tccaggatgg cagcaatggc gctgtgcggc ctcaccaggc ccacgaggag cagcagcgt  
2461 cggcccggag cagcaggaag gccctctgtg ggagcgcccg ccgtctgctc cggggtgggt  
2521 cagtcactgc ttgttgacat caacatggca attgcactca tgtggactgg gaccgtgca

2581 gctgccgtgt gggtagtcg ggtgccagga caatgaaata ctccagcacg tgtggctgac  
 2641 gaatttgttt ctacagaaat aacagctggg gacaactgcg gtgatgatgt aaaaaccttc  
 2701 ccataaaatg taagaaaagc tgatgaggct ggtgacgttc agcctttgtc aataaacctg  
 2761 tcattgtcgg atcctt

## 5 SEQ ID NO: 136

Amino acid sequence of human GPR30 encoded by the DNA sequence shown in SEQ ID NO: 135.

MDVTSQARGVGLEMPGTAQPAAPNTTSPELNLSHPLLGTALANGTSELSEHQYVIGLF  
 10 LSCLYTIFLPIGFVGNILILVNVISFREKMTIPDLYFINLAVADLILVADSLIEVFNH  
 ERYDYIAVLCTFMSLFLQVNMYSVFFLTWMSFDRIYALARAMRCSLFRKHHARLSCGL  
 IWMASVSATLVPFTAVHLQHTDEACPCFADVREVQWLEVTLGFIIVPFAIIGLCYSLIVRV  
 LVRHRHRLRPRRQKALRMILAVVLVFFVCWLPENVFISVHLLQRTQPGAAPCKQSFH  
 AHPLTGHIVNLAAPFNSCLNPLIYSFLGETFRDKLRLYIEQKTNLPALNRFCHAALKAVI  
 PDSTEQSDVRFSSAV

## 15 SEQ ID NO: 137

gi|38081363|ref|XM\_355659.1| Mus musculus G protein-coupled receptor 30 (Gpr30), mRNA

1 ataaaggagg cgctgtgccg agggggccag acgctgctgg acggccacag gcatccatcc  
 61 ccaggcatcg ggccgggtgct tctgttcctc tccctgctgg tccctgctgg gcaccgtccc  
 120 121 caaagtgtcg caagtccagg gtccatccct ggagcaagct ccaggagcac ctccagcaga  
 181 tggcctggta acgccacggc acagatcagg acaccaaca gaaaatcaga aggacactaa  
 241 gtctgatcgt tagattaaca gacgacgcat ctggaccaag gacagaagcc aggggtgcat  
 301 ttctgccatg caccaccaa aacagctgat cagatctagg gagaaagcca tccaaggact  
 361 ctgctccctc taagctgctg gaattgtggc caagcctcaa cactcacaca ctctgggtgc  
 421 ccagaagggtg agcaggcagc aggtgtgcct gccagcacc agccagaca tcagacaccc  
 481 tgtccaccct tctggttttc tgagactaac aggtccctcag gacgattctt cctgcctcac  
 541 aaatgcctgg tttatcttct tttgtgaaga tggagctgtc acataaaaca gctttctgtg  
 601 accctttcag caaatcctga aaactgccga gggaagccat ggatgcgact actccagccc  
 661 aaactgttgg ggtggagatc tacctaggtc cgtgtggcc agcccttcc aacagacccc  
 721 ctctggccct caacttgctc ctggcactgc gggaagatgc cccggggaac ctcatgggg  
 781 acctctctga gcatcagcag tacgtgattg ccctcttctc ctctgctc tacaccatct  
 841 tctcttttcc tattggcttt gtgggcaaca tctctatcct ggtggtgaac atcagcttcc  
 901 gggagaagat gaccatccca gacctgtact tcatcaacct ggccggcggc gacctcatcc  
 961 tgggtggtga ctccctgatt gaggtgttca acctggacga gcagtactac gacatcgag  
 1021 tgctctgcac ctccatgtcc ctcttcctgc agatcaacat gtacagcagc gtcttcttcc  
 1081 tcacctggat gagcttcgac aggtacctag cgtggccaa ggccatgcg tgtggcctct  
 1141 tccgcaccaa gcaccacgca cggctcagct gtggcctcat ctggatggcc tcagtgtccg  
 1201 ccacgtggtg gcccttcaca gcgtgcacc tgcggcacac ggaggaggcc tgcttctgct  
 1261 ttgtgatgtg caggagggtg cagtggctgg aggtcacact gggcttcac atgcccttcg  
 1321 ccacatttgg cctctgctac tccctcatcg tgcgagccct catccgggcc cacaggcacc  
 1381 gcggcctgcg cccacgcagg cagaaagccc tgaggatgat ctccgcagtg gtccttgttt  
 1441 tcttcatctg ctggctgccc gagaacgtct tcatcagtg ccacctactg cagtggacgc  
 1501 agccagggga cactccctgc aagcagctct tccgtcacgc ctacccttg acaggccaca  
 1561 tagtcaacct tgcagccttc tccaacagct gctgaatcc cctcatctac agcttctggtg  
 1621 gagagacctt cagggacaag ctcaggctct atgtggagca gaagacgagc ctgcccgtctc  
 1681 tgaaccgctt ctgccatgcc acgtcaagg ccgtcatctc agacagcaca gagcagtcag  
 1741 aggtcagggt cagcagtgct gtgtgagagg aaaaggtcag gggcgaggc tgggtgctcag  
 1801 gacttgacac cacttagcac aggtgggtgag tgggctaagc tatgtcatac tctcaaacc  
 1861 cagtggcttg gggaagacgt cacattgccc ggtcatctct ggagctgctg gcatccttcc  
 1921 tgactgtcca gctcatggat gctgccatcc agattcaagg tcccaaggca gcggggcacc  
 1981 tgactattgac ctctgacctc aaagggcacc agcgggcctc gctgcttggc tttctttcca  
 2041 tagcctacgt tcccagaaca caagtctgct gttgatacga ggacaggcca tgctatggga

2101 gcaccatggtt acatgcctgc tacgtggagg agtctagaga cagactttat gtaccagacc  
 2161 caaactggct accttccctt tgcttgatgt gtgtaactga ccatgtatac accgtccagt  
 2221 gcagccagag ccttcttccct gtcttccaga aggtctgtgag gtcacccag atgccactcc  
 2281 taactcctga gtgaacagcg tgtctgactg agaaaggccc ttttaacaaaa cgccttcctg  
 5 2341 ctctgggatg ctctctcac aaagtttgtt taaaaagggtg tttgcccttc cgtgaagggtg  
 2401 gaaggagact ggggtgctgct gtgcaggctg gtgggatgcc gccataagat gtgtggtaga  
 2461 aggacttacc accacagaaa atcatactgg gaacagcgag ctgtaaattg atctcattaa  
 2521 aacgt

SEQ ID NO: 138

- 10 Amino acid sequence of mouse GPR30 encoded by the DNA sequence shown in SEQ ID NO: 137.

MDATTPAQTVGVEIYLGPVWPAPSNSTPLALNLSLALREDAPGNLTGDLSEHQYVIALF  
 LSCLYTIFLFPPIGVGNILILVVNISFREKMTIPDLYFINLAAADLILVADSLIEVFNL  
 EQYYDIAVLCTFMSLFLQINMYSSVFFLTWMSFDRLALAKAMRCGLFRTKHHARLSCGL  
 15 IWMAVSATLVPFTAVHLRHTTEACFCFADVREVQWLEVTLGFI MPFAIIGLCYSLIVRA  
 LIRAHRRHRLRPRRQKALRMIFAVVLVFFICWLPENVFISVHLLQWTPGDT PCKQSPRH  
 AYPLTGHIVNLAASFNSCLNPLIYSFLGETFRDKLRLYVEQKTSLPALNRFCHATLKA  
 VPDSTEQSEVRFSSAV

SEQ ID NO: 139

- 20 gi|19424261|ref|NM\_133573.1| Rattus norvegicus G protein-coupled receptor 30 (Gpr30), mRNA

1 ttctgtgacc ctttcagcaa gtccctgaaag cttctacggg aagccatggc tgcaactact  
 61 ccagcacaaag atgttgccgt agagatctac ctagggtccc tgtggccagc cccttccaac  
 121 agcacccttc tggccctcaa cctgtccctg gcgctgcggg aagatgcccc ggggaacctc  
 25 181 actggggacc tctctgaaca tcagcaatat gtgatcgctc tcttctcttc ctgctctac  
 241 accatcttcc tcttcccat cggctttgtg ggcaacatcc tcatcttggg ggtgaacatc  
 301 agcttccggg agaagatgac tatccagac ctgtacttca tcaacctggc agcggctgac  
 361 ctcatcctgg tggccgactc cctgatcgag gtgttcaacc tggacgagca gtattacgat  
 421 atcgccgtgc tctgcacctt catgtccctc ttctgcaga tcaaatgta cagcagctc  
 30 481 ttcttctctc cctggatgag cttcgacagg tacctggcgc tggccaaagc catgcgctgt  
 541 ggcctcttcc gcaccaagca ccacgcggc ctcagctgtg gcctcatctg gatggcctca  
 601 gtgtccgcca cgctgggtgcc cttcacggc gtgcatctgc ggcacacoga ggaggcctgc  
 661 ttctgctttg ccgatgtcag ggaggtgcag tggctggagg tcacgctggg cttcattgtg  
 721 cccttcgcca tcatcgccct gtgtattcc ctcacgtgc gggccctcat ccgggcccac  
 35 781 aggcacgtg gcctgcgccc acgcaggcag aaagccctga ggatgatctt cgcagtggc  
 841 cttgtcttct tcatctgctg gctgccggag aacgtcttca tcagcgtcca cctactgcag  
 901 tgggcgcagc caggggacac tccctgcaag cagtcttcc gtcacgcta ccccttgaca  
 961 ggccacatag tcaacctggc agccttctcc aacagctgcc tgagtccct catctatagc  
 1021 ttcctgggag agaccttcag ggacaagctc aggtctgatg tggcgagaa gacgagcctg  
 40 1081 ccagctctca accgcttctg ccatgccacg ctcaaggcag tcataccaga cagcacggag  
 1141 cagtcagatg tcaagttcag cagtgtgta tgagaggtac ctctagagg aaaacggaca  
 1201 ggggagcagg cgtgcccagg agctgcacac tctagcacag gtggtgggag agctgagcca  
 1261 tgtcatactc taaacccc

SEQ ID NO: 140

- 45 Amino acid sequence of rat GPR30 encoded by the DNA sequence shown in SEQ ID NO: 139.

MAATTPAQDVGVEIYLGPVWPAPSNSTPLALNLSLALREDAPGNLTGDLSEHQYVIALF  
 LSCLYTIFLFPPIGVGNILILVVNISFREKMTIPDLYFINLAAADLILVADSLIEVFNL

EQYYDIAVLCTFMSLFLQINMYSSVFFLTWMSFDRLALAKAMRCGLFRTKHARLSCGL  
 IWMAVSATLVPFTAVHLRHTTEACFCFADVREVQWLEVTLGFI VPF AI IGLCYSLIVRA  
 LIRAHRRHGLRPRRQKALRMIFAVVLVFFICWL PENVFISVHLLQWAQPGDTPCKQSF RH  
 AYPLTGHIVNLAAFSNSCLSP LIYSFLGETFRDKLRLYVAQKTS L PALNRFCHATLKAVI  
 5 PDSTEQSDVKFSSAV

SEQ ID NO: 141

gi|33695103|ref|NM\_003608.2| Homo sapiens G protein-coupled receptor 65 (GPR65), mRNA

```

10      1 ttcttgactt gatgcaggca cagatttata aagctcctca gtcaacaaac acatcaccgg
      61 aagaaacatg gaaggaaaagg aatttttaaaa ggaaatacca atctctgtgc aaacaaagcc
      121 ttgtatatcc atgtttgcac caatctactg tgagatttat gaagaaaaac aaattgcgga
      181 caactctcta tgtacactta caaatgcctc agttgatgct tgtgggctgt ttgtcagcgt
      241 tctgtgataa tgaacacatg gacttctgtt tattaatttc agttgacccc ttagccaat
      301 tgccaggagc ctggattttt acttccaact gctgatatac gtgtaaaaat tgatctacat
15      361 ccacccttta aaagcattga tgaattaatt agaactttag acaacaaaga aaaattgaaa
      421 aagaattctc agtaaaagcg aattcgatgt tcaaaacaaa ctacaaagag acaagacttc
      481 tctgtttact ttctaagaac taatataatt gctaccttaa aaaggaaaaa atgaacagca
      541 catgtattga agaacagcat gacctggatc actatttgtt tcccatgtgt tacatctttg
      601 tgattatagt cagcattcca gccaatattg gatctctgtg tgtgtctttc ctgcaagcaa
20      661 agaaggaaag tgaactagga atttacctct tcagtttgtc actatcagat ttactctatg
      721 cattaaactct ccctttatgg attgattata cttggaataa agacaactgg actttctctc
      781 ctgccttgtg caaaggaggat gcttttctca tgtacatgaa tttttacagc agcacagcat
      841 tcctcacctg cattgccggt gatcggtatt tggctgttgt ctaccctttg aagttttttt
      901 tcctaaaggac aagaagattt gcactcatgg tcagcctgtc catctggata ttggaaacca
25      961 tcttcaatgc tgtcatgttg tgggaagatg aaacagtgtg tgaatattgc gatgccgaaa
      1021 agtctaattt tactttatgc tatgacaaat accctttaga gaaatggcaa atcaacctca
      1081 acttggttcag gacgtgtaca ggctatgcaa tacctttggg caccatcctg atctgcaacc
      1141 ggaaagtcta ccaagctgtg cggcacaata aagccacgga aaacaaggaa aagaagagaa
      1201 tcataaaact acttgctcagc atcacagtta cttttgtctt atgctttact ccctttcatg
30      1261 tgatgttgct gattcgctgc atttttagagc atgctgtgaa cttcgaagac cacagcaatt
      1321 ctgggaagcg aacttacaca atgtatgaa tcacggttgc attaacaagt ttaaattgtg
      1381 ttgctgtacc aattctgtac tgttttgtaa ccgaaacagg aagatatgat atgtggaata
      1441 tattaataatt ctgcactggg aggtgtaata catcacaaag acaaaagaaa cgcatacttt
      1501 ctgtgtctac aaaagatact atggaattag aggtccttga gtagaaccac ggatgttttg
35      1561 aagggaaggg aagttaagt tatgcattat tatatcatca agattacatt ttgaaaagga
      1621 aatctagcat gtgaggggac taagtgttct cagagtgtat ttttaacca gtccaataaa
      1681 aatatcttaa aactgcattg tacagctccc tcctgcggtt ttattaaatg atgtatatta
      1741 aacaaagatc aataaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa

```

SEQ ID NO: 142

40 Amino acid sequence of human GPR65 encoded by the DNA sequence shown in SEQ ID NO: 141.

```

      MNSTCIEBQHDLDHYLFPIVYIFVIIVSIPANIGSLCVSFLQAKKESELGIYLFSLSLSD
      LLYALTLPWLIDYTWKDNWTFSPALCKGSAFLMYMNFYSSTAFLTCIAVDRLAVVYPL
      KFFFLRTRRFALMVSLSIWILETIFNAVMLWEDETVEYCDAEKSNFTLCYDKYPLEKWQ
45      INLNLFRCTGYAIPLVTLICNRKVYQAVRHKNATENKEKKRIKLLVSITVTVLCFT
      PFHVMLLIRCILEHAVNFEDHSNSGKRITYMYRITVALTSLNCVADPILYCFVTETGRYD
      MWNILKFTGRCNTSQRQRKILSVSTKDTMELEVLE

```

SEQ ID NO: 143



gi|1103872|gb|U39827.1|MMU39827 Mus musculus putative G protein-coupled receptor TDAG8 (TDAG8) mRNA, complete cds

```

1  gatacaagca gattttgccag cctcctcagt caagagaagc atccctccag aaacagggaa
5  61  acatgacact tttgaaagaa tgccaaacgg cgtgaaaata aaaacagagc attcccattt
121 gcaccgacca atctccaatc tctgtgaaga ttcaaaaggg caagcaagag gcggtgaccg
181 ttcacgaaag ctaaaatccc atgctattga acatgaagac ttctgatgct taaatctcat
241 taactgcttt aagtcactcc caggagcttg gatcccaact tctagcagta atagtctgtg
301 taaaaaaaaa aaaaaaatca gtctacaacc actctctaaa tgcattggatg aactcatcag
361 aacatcaaaa cccaaggaaa ccctaagaga gaagaattct aataaaaaga attttacatt
10 421 gaaaacttac aaggcaaggt ccctttccct gctgacagcc taagaagtga tgtaactgcc
481 actgtgaaga ccatggcgat gaacagcatg tgcattgaag agcagcgcca cctcgaacac
541 tattttgttcc cgggtggtcta cataattgtg ttatatagta gcgtcccagc caacatcgga
601 tctttatgcy tctcctttct gcaagcgaag aaggaaaaatg agctagggat ttacctcttc
661 agtctgtccc tgtcagacct gctgtatgcy ctgactctgc ccctctggat caattacact
15 721 tggaaataaaa acaactggac ttctctctcc accttgtgca aaggaaagcgt ttcttcacc
781 tacatgaact tttacagcag cacggcgctt ctcacttgca ttgccctgga ccgctattta
841 gcagtcgtct accctctgaa gttttccttc ctaagaacga gaagattcgc gtttattacc
901 agcctctcca tctggatatt agagtccttc tttaactcta tgcttctgtg gaaagatgaa
961 acgagtgttg aatattgtga ctcggaacaa tctaatttca ctctctgcta tgacaaatac
20 1021 cctctggaga aatggcagat aaacctcaac ctgtttcgga cgtgcatggg ctacgcaata
1081 cccttgatca ccatcatgat ctgcaacctt aaagtctacc gagctgtgcy gcacaaccaa
1141 gccacggaaa acagcgagaa gagaaggatc ataaagttgc ttgctagcat cacgttgact
1201 ttgctcttat gctttacccc cttccacgtg atgggtgctc tccgctgcyt tttagagcgc
1261 gacatgaacg tcaatgacaa gtctggatgg cagacgttta cgggtgtacag agtcacagta
25 1321 gccctgacga gtctaaactg tgttgccgat cccattctgt actgctttgt gactgagacg
1381 gggagagctg atatgtggaa catattaaaa ttgtgtacta ggaaacacaa tagacaccaa
1441 gggaaaaaaa gggacatact ttctgtgtcc acaagagatg ctgtagaatt agagattata
1501 gactaagagc tggaggcagg ttaagttaca tgggtattat taatgaaact tacatttttg
1561 aaaagaatac tggcatagta gaaccagatg gaaatagttt gaaggtacat tgtatgactc
30 1621 ctatgtttggc tttattaagt aaggtataga aatgtattat cttgtatgta ttctaattgac
1681 taggcacatc tgttttagta ccaattctct ttgcctctat gttataaccc ctaagaagca
1741 cgcgggactg ttctgtctta aatcagtggc cattctatct gactactatg actttttgtt
1801 gttgttctgc tttgggtttt cagtctgcct gcacagtcct tctcctctgt atacgtctgt
1861 cttcaacaaa tgtaaggact aaatacccct cccgatcaca tccattatca aggatttgaa
35 1921 gccactccat gtactgggtt ataaaagaaa tgttctcatg aactttcatg aagtttacat
1981 acctttgggg atctagtcac cgagtcacat aaagtaaaag taaatggatt ttgtan

```

SEQ ID NO: 144

40 Amino acid sequence of mouse GPR65 encoded by the DNA sequence shown in SEQ ID NO: 143.

```

MAMNSMCIEEQHHLEHYLFPVYIIVFIVSVPANIGSLCVSFLQAKKENELGIYLFSLSL
SDLLYALTLPLWINYTWKDNWTFSPTLCKGSVPFTYMNFYSSAFLTCIALDRYLAVVY
PLKFSFLRTRRFAFITSLSIWILESFFNSMLLWKDETSVEYCDSDKSNFTLCYDKYPLEK
45 WQINLNLFRCTMGYAIPLITIMICNHKVYRAVRHNQATENSEKRRIKLLASITLTFVLC
FTPFHVMVLIRCVLERDMNVNDKSGWQTFYRVTVALTSLNCVADPILYCFVTETGRAD
MWNILKLCTRKHNHQGQKRDILSVSTRDAVELEIID

```

SEQ ID NO: 145

gi|27667219|ref|XM\_234367.1| Rattus norvegicus similar to Gpcr25 protein (LOC299242), mRNA

```

50 1  atgacgatga acagcacatg tgttgaagaa cagcatgacc tggaccacta tttgttccca
61  gtgggtctaca tatttgtgtt tatagtcagc gtcccagcca acatcggtctc tttatgtgta

```

```

121 tcttttctgc aagcgaagaa ggaaaatgaa ctaggagttt acctcttcag tctgtcactg
181 tcagatctgc tgtatgcgtt gacgctgcct ctctggataa actacacttg gaataaagac
241 aactggacgc tctctccac cttgtgcaaa ggaagcgttt tcttcaccta cctgaacttt
301 tacagcagca cggccttcct cacttgcat gccttgacc gctatttagc agtcgtctac
5 361 cctctaaggt tttcttttct aagaacaaga agatttgcat ttatcaccag cctgtctatc
421 tggatattgg aatccgtctt taactccatc cttctgtggg aagacgaaac aagtgtcgaa
481 tactgtgatg cagagaaatc taatttcact ctctgctatg acaataacc tctggagaag
541 tggcagataa accttaacct gttcaggact ttcttgggct acgcaatacc cttggtcacc
601 atcatgatct gcaaccataa agtctaccgc gccgtgaagc acaaccaagc tacggaaaac
10 661 aacgaaagga gacggatcat aaagctgctt gccagcatca cgctgacttt catcttatgc
721 tttacccttt tccacgtgat ggtgctcatc cgctgcattt tagagcgcaa cgtgaacttc
781 aataacaaga ctggaaagca gacatttcac gtgtacagaa tcacagtggc cctgacgagt
841 ctaaaactgtg tcgccgatcc aattctctac tgtttcgtga cggaaaccgg gagatccgat
901 atgtggaaca tattaagaat gtgtgctagg aagcgcaata gatcccgagg aaaaaaaga
15 961 gacatacttt ctgtgtccac aagagatacc atagaactgg agattataga ataa

```

SEQ ID NO: 146

Amino acid sequence of rat GPR65 encoded by the DNA sequence shown in SEQ ID NO: 145.

```

20 MTMNSTCVVEQHDLDHYLFPPVYIFVPIVSV PANIGSLCVSFLQAKKENELGVYLFSLSL
SDLLYALTLP LWINYTWNKDNWTLSP TLCKGSVF FTYLNFYSSTAF LTCIALDRYLAVVY
PLRFSFLRRRFAFITSL SIWILESVFNSILLWEDETSVEYCD AEKSNFTLCYDKYPLEK
WQINLNLFR TFLGYAIPLV T IMICNHKVYRAVKHNQATENNERRRI I KLLASITLTFILC
25 PTPFHVMLIR CILERNVNFNNK TGKQTF TVYRITVALTSLNCVADPILYCFVTETGRSD
MWNILKVCARKNRNRSRGQKRDILSVSTRDTIELEIE

```

SEQ ID NO: 147

gi|40254430|ref|NM\_000867.2| Homo sapiens 5-hydroxytryptamine (serotonin) receptor 2B (HTR2B), mRNA

```

1 gggggtat ttttactgc tttcaaccgc ctgtgctgga ggctcagaat aagtcaatgg
30 61 gagggaggatt tcagtcacag cagcaagcaa gtctagtga cagataagat gacatgctca
121 gcaaaataac aacgaaacca gagggggaac tctctggcat gcaagttcaa acacgactct
181 acaactacgg cagaaaaaga gagagagaga aactaaaaat atatatatat cctatttttt
241 tcacagctat cagtttcttt cactgagctt tcctaaat t aagcctctag aaaataataa
301 atacttggat atcttaccta caaacatgga cagatgtgtg tatgcgctca ttttagagaa
35 361 ctgtaatttt ttttttaaa ggaaggtgtc aactttggct tttgagtgtt tggcatgggt
421 acaatgcctt aaaaaaacag atgagcagct tagctactaa ccatgctgac cactgttcgg
481 aacgggattg aatcacagaa aaacagcaa tggctctctc ttacagagtg tctgaacttc
541 aaagcacaat tcctgagcac attttgaga gcacctttgt tcacgttatc tcttctaact
601 ggtctggatt acagacagaa tcaataccag aggaaatgaa acagattgtt gaggaacagg
40 661 gaaataaact gcactgggca gctctctga tactcatggg gataataccc acaattgggt
721 gaaataccct tgttattctg gctgtttcac tggagaagaa gctgcagtat gctactaatt
781 actttctaata gtccttggcg gtggctgatt tgctggttgg attgtttgtg atgccaatg
841 cctcttgac aataatgttt gaggtatgtt ggccctccc acttgttcta tgcctgcct
901 gggtatttct tgacgttctc tttcaaccg catccatcat gcatctctgt gccatttcag
45 961 tggatcggtta catagccatc aaaaagccaa tccaggccaa tcaatataac tcacgggcta
1021 cagcattcat caagattaca gtggtgtggt taatttcaat aggcattgcc attccagtc
1081 ctattaaagg gatagagact gatgtggaca acccaaacia tatcacttgt gtgctgacaa
1141 aggaacgttt tggcgatttc atgctctttg gctcactggc tgccttcttc acactcttg
1201 caattatgat tgtcacctac tttctcacta tccatgcttt acagaagaag gcttaacttag
50 1261 tcaaaaacaa gccacctcaa cgcctaacta ggttgactgt gtctacagt tcccaaagg
1321 atgaaacacc ttgctcgtca cgggaaagg tggcaatgct ggatgggtct cgaaaggaca
1381 aggctctgcc caactcaggt gatgaaacac ttatgcgaag aacatccaca attgggaaaa

```

1441 agtcagtgcg gaccatttcc aacgaacaga gagcctcaaa ggctcctaggg attgtgtttt  
 1501 tcctcttttt gcttatgtgg tgcctcttct ttattacaaa tataacttta gttttatgtg  
 1561 attcctgtaa ccaactact ctccaaatgc tcctggagat attgtgtgg ataggctatg  
 1621 tttcctcagg agtgaatcct ttggctaca ccctcttcaa taagacattt cgggatgcac  
 5 1681 ttggccgata tatcacctgc aattaccggg ccacaaaagtc agtaaaaaact ctacagaaac  
 1741 gctccagtaa gatctacttc cggaaatcaa tggcagagaa ctctaagttt ttcaagaaac  
 1801 atggaattcg aaatgggatt aaccctgcc a tgtaccagag tccaatgagg ctccgaagtt  
 1861 caaccattca gtcttcatca atcattctac tagatacgtc tctcctcact gaaaatgaag  
 1921 gtgacaaaac tgaagagcga gttagttatg tatagcagaa ctggcagttg tcatcaaaac  
 10 1981 taatgatgag taagatgatg aatgagatgt aaatgtgcca agaataatatt atataaagaa  
 2041 ttttatgtca tatatcaaat catctcttta acctaagatg taagtattaa gaatatctaa  
 2101 ttttctaatt ttggacaaga ttattccatg aggaaaataa ttttatatag ctacaaatga  
 2161 aaacaatcca gcaactctgg taaattttta ggtattcgaa tgaaataaag tcaaatcaat  
 2221 aaatttcagg ccaaaaaaaa aaaaaaaaaa aaaaaaaaaa

## 15 SEQ ID NO: 148

Amino acid sequence of human HTR2B encoded by the DNA sequence shown in SEQ ID NO: 147.

20 MALSYRVSELQSTIPEHILQSTFVHVISSNWSGLQTESIPEEMKQIVEEQGNKLHWAALL  
 ILMVIIPTIGGNTLVILAVSLEKKLQYATNYFLMSLAVADLLVGLFVMPIALLTIMFEAM  
 WPLPLVLCPAWFLDVLFTASIMHLCAISVDYIAIKKPIQANQYNSRATAFIKITVVW  
 LISIGIAIPVPIKGIETDNDPNITCVLTKERFGDFMLFGSLAAPTPLAIMIVTYFLT  
 IHALQKKAYLVKNKPPQRLTWLTVSTVFQRDETPCSSPEKVMALDGSRKDKALPNSGDET  
 LMRRTSTIGKKSQVQTSNEQRASKVLGIVFFLLMWCFFITNITLVLCDSCNQTTLMQ  
 25 LLEIFVWIGYVSSGVNPLVYTLFNKTFRDAFGRYITCNRYRATKSVKTLRKRSSKIYFRNP  
 MAENSKFFKKHGIRNGINPAMYQSPMRLRSSTIQSSSIILLDTLLLLENEDKTEERVS  
 V

## SEQ ID NO: 149

gi|6680322|ref|NM\_008311.1| Mus musculus 5-hydroxytryptamine (serotonin) receptor 2B (Htr2b), mRNA

30 1 actgtctgga actggactga gtcacaaaaa ggcgaaatggc ttcattcttat aaaatgtctg  
 61 aacaaagcac aacttctgag cacattttac agaagacatg tgatcacctg atcctgacta  
 121 accgttcttg attagagaca gactcagtag cagaggaaat gaagcagact gtggaggagc  
 181 aggggcatac agtgcactgg gcagctctcc tgatactcgc ggtgataata cccaccattg  
 241 gtgggaacat ccttgtgatt ctggctggtg cactggagaa aaggctgcag tacgtacca  
 35 301 actacttttt aatgtccttg gcgatagcag atttgcctgg tggattgttt gtgatgccga  
 361 ttgccctctt gacaatcatg tttgaggcta tatggccctt cccactggcc ctgtgtcctg  
 421 cctgggttatt cctcgatggt ctcttttcaa ctgcctccat catgcatctc tgtgccattt  
 481 ccctggaccg ctatatagcc atcaaaaagc caattcaggc caatcagtgc aacacccggg  
 541 ctactgcatt catcaagatt acagtgggat ggtaatttc aataggcatc gccatcccag  
 40 601 tccctattaa aggaatcgag actgatgtga ttaatccaca caatgtcacc tgtgagctga  
 661 caaaggaccg ctttggcagt tttatggctt ttgggtcact ggctgcttcc ttcgtacctc  
 721 tcaccatcat gtagtgcact tactttctca ccattcacac ttacagaag aaagcttact  
 781 tggtaaaaaa taagccacct caacgcctaa cacggtggac tgtgccaca gtttctctaa  
 841 gggaagactc atccttttca tcaccagaaa aggtggcaat gctggatggg tctcacaggg  
 45 901 ataaaattct acctaactca agtgatgaga cacttatgag aagaatgtcc tcagtggaa  
 961 aaagatcagc ccaaacattt tctaagagc agagagcctc gaaggccctt ggagtcgtgt  
 1021 ttttcccttt tctgcttatg tgggtgccct ttttattac aaatctaact ttagctctgt  
 1081 gtgattcctg caatcagacc actctcaaaa cactcctgga gatatttggt tggataggct  
 1141 acgtttcctc gggggatgaat cctctgatct atacactctt caataagaca tttcgggaag  
 50 1201 catttggcag gtacatcacc tgcaattacc gagccacaaa gtcagtataa gcacttagga  
 1261 agttttccag tacactttgt tttgggaatt caatggtaga aaactctaaa ttttcacaa  
 1321 aacatggaat tcgaaatggg atcaaccctg ccatgtacca gagcccaatg aggtccgat

1381 gttcaaccat tcagtcctca tcaatcatcc tcctcgatac ccttctcact gaaaacgatg  
 1441 gcgacaaagc ggaagagcag gtcagctaca tattgcagga acgggccggc ctcattctga  
 1501 gagaggggtga tgagcaggac gcacgcgcac catggcaggt tcaagagtga

SEQ ID NO: 150

- 5 Amino acid sequence of mouse HTR2B encoded by the DNA sequence shown in SEQ ID NO: 149.

MASSYKMQSTTSEHILQKTCDHILITNRSGLTDSVABEMKQTVGQGHVHWAALLI  
 LAVIIPITIGNILVILAVALEKRLQYATNYFLMSLAIDLLVGLFVMPIALLTIMFEAIW  
 10 PLPLALCPAWLFLDVLFPSTASIMHLCAISLDRIYIAIKKPIQANQCNRATAFIKITVVWL  
 ISIGIAIPVPIKGIETDVINPHNVTCELTQRFGSFMVFGSLAFAFFVPLTIMVVTYFLTI  
 HTLQKKAYLVKNKPPQRLTRWTVPTVFLREDSSFSPEKVAMLDGSHRDKILPNSSDETL  
 MRRMSSVGKRSQATISNEQRASKALGVVFFLLMWCPPFITNLTLALCDSCNQTTLKT  
 LEIFVWIGYVSSGVNPLIYTLFNKTFREAFGRYITCNRYRATKSVKALRKFSSTLCFGNSM  
 15 VENSKEFTKGIKIRNGINPAMYQSPMLRRCSTIQSSSIILLDTLLTENDGDKAEQVSYIL  
 QERAGLILREGDEQDARAPWQVQE

SEQ ID NO: 151

gi|8393585|ref|NM\_017250.1| Rattus norvegicus 5-hydroxytryptamine (serotonin) receptor 2B (Htr2b), mRNA

1 ctgaaatcta agcctctaga aggactagaa tctggatgtc ttacctgcaa acatggacag  
 20 61 atatgtacac agtcccatct tggagaacct gaatcttttt agaagaaaga aggccacctt  
 121 ggctgggagt gtctggagga taccatgctt tgcaaaagca gatgacctgc tagcaactga  
 181 ccatgctgac cactgtcttg aactggactg agtcacagaa aggcgaatgg cttcatctta  
 241 taaaatgtct gaacaaagca caatttctga gcacattttg cagaaaacat gtgatcacct  
 301 gatcttgact gaccgttctg gattaaaggc agaatcagca gcagaggaaa tgaagcagac  
 25 361 tgccgagaac cagggaata cagtgcactg ggcagctctc ctgatcttcg cggtaataat  
 421 ccccaccatt ggcgggaaca tcctgggtat tctggctgtt tctactggaga aaaggctgca  
 481 gtacgtacc aactactttc taatgtcctt ggcgggtggc gatttgctgg ttggattggt  
 541 tgtgatgccg attgctctct taacaatcat gtttgaggct acatggcccc tcccactggc  
 601 cctgtgtcct gcttggttat ccttgatgt tctcttttca actgcctcca tcatgcatct  
 30 661 ctgtgccatt tccctggatc gctatatag catcaaaaag ccaattcagg ccaatcagt  
 721 caattcccg actactgcat tctgaagat tacgggtgga tgggttaattt caataggcat  
 781 cgccatcca gtccctatta aaggaataga ggctgatgtg gtcaacgcac acaatcac  
 841 ctgtgagctg acaaaggacc gctttggcag tttcatgctc tttgggtcac tggctgctt  
 901 ctttgcacct ctcaccatca tgatagtcac ctactttctc accattcacg ctttgcggaa  
 35 961 gaaagcttac ttggtcagaa acaggccacc tcaacgccta acacgggtgga ctgtgtccac  
 1021 agttctccaa agggaagact catccttttc atcaccagaa aagatggtga tgctggatgg  
 1081 ctctcacaag gataaaattc tacttaattc aactgatgag acactgatga gaagaatgtc  
 1141 ctccagcagga aaaaaaccag cccagaccat ttctaataga cagagagcct caaaggctcct  
 1201 tgggaattgtg tttctcttct ttctgcttat gtggtgcccc tttttcatta caaacgtaac  
 40 1261 ttttagctctg tgtgattcct gcaaccagac tactctcaaa acactcctgc agatatttgt  
 1321 gtgggtaggc tacgtttcct cgggagtgaa tcctttgatc tataacctct tcaataagac  
 1381 atttcgggaa gcatttggca ggtacatcac ctgcaattac caggccacaa agtcagtaaa  
 1441 agtgcttaga aagtgttcta gtacactcta ttttgggaat tcaatggtag aaaactctaa  
 1501 atttttcaca aaacatggaa ttcgaaatgg gatcaaccct gccatgtacc agagccagc  
 45 1561 aaggctccga agttcaacca ttcagtcttc atccatcatt ctctcaata catttctcac  
 1621 tgaacacgat ggtgacaaag tagaagacca agtcagctac atatagtgga atggggcagc  
 1681 cctcatctga ctgagggagg ggatgaggag gacgcaagca aaccaaggaa aaggcaagag  
 1741 tgaagcacta aggttgtcca gtttccttat ctaaacaaac tcaacgcacg ggtatagtag  
 1801 ttccgtatgg ctacaaacaa aagcattccc tactctggtt tcaaatgga acaaaattaa  
 50 1861 ataagtggat atacttcagt ctttaaaaag aaaagaaggg gttggggatt tagctcagtg  
 1921 gtagagagtt tgcctagtaa gtgcaaggcc ctgggttcgg ttctcagctc cagaaaaaaa  
 1981 aaaaataaaa aaaaaaata aataaaaaaa aaaaaaaaaa a

## SEQ ID NO: 152

Amino acid sequence of rat HTR2B encoded by the DNA sequence shown in SEQ ID NO: 151.

5 MASSYKMSEQSTISEHILQKTCDHILITDRSGLKAESAAEEMKQTAENQNTVHWAALLI  
FAVIIPTIGGNILVILAVSLEKRLQYATNYFLMSLAVADLLVGLFVMPIALLTIMFEATW  
PLPLALCPAWLFLDVLSTASIMHLCAISLDRIYIAIKKPIQANQCNSRTTAFVKITVVWL  
ISIGIAIPVPIKGIADVVNAHNITCELTDRFGSFLMFGSLAAFFAPLTIMIVTYFLTI  
HALRKKAYLVNRNPPQRLTRWTVSTVLQREDSSFSSPEKVMMLDGSCHKILPNSTDETL  
10 MRRMSSAGKKPAQTISNEQRASKVLGIVFLFFLLMWCFFFITNVTLALCDSCNQTTLKTIL  
LQIFVWVGYYVSSGVNPLIYTLFNKTFREAFGRYITCNQATKSVKVLKCSSTLYFGNSM  
VENSKFFTKHGIRNGINPAMYQSPVRLRSSTIQSSSIILLNTFLTENDGDKVEDQVSYI

## SEQ ID NO: 153

gi|4557885|ref|NM\_000211.1| Homo sapiens integrin, beta 2 (antigen CD18 (p95),  
lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)  
15 (ITGB2), mRNA

1 cagggcagac tggtagcaaa gccccacgc ccagccagga gcaccgccgc ggactccagc  
61 acaccgaggg acatgctggg cctgcgcccc ccactgctcg ccctgggtggg gctgctctcc  
121 ctcggtgctg tctctctca ggagtgcacg aagttcaagg tcagcagctg ccgggaatgc  
181 atcgagtcgg gggccggctg cactggtgac cagaagctga acttcacagg gccgggggat  
20 241 cctgactcca ttcgctgcga caccgggcca cagctgctca tgaggggctg tgcggctgac  
301 gacatcatgg accccacaag cctcgctgaa acccaggaag accacaatgg gggccagaag  
361 cagctgtccc caaaaaagt gacgctttac ctgcgaccag gccaggcagc agcgttcaac  
421 gtgaccttcc ggcgggcca gggctacccc atcgacctgt actatctgat ggacctctcc  
481 tactccatgc ttgatgacct caggaatgtc aagaagctag gtggcgacct gctccgggcc  
25 541 ctcaacgaga tcaccgagtc cggcgcatt ggcttcgggt ccttcgtgga caagaccgtg  
601 ctgccgttcg tgaacacgca ccctgataag ctgcgaaacc catgccccaa caaggagaaa  
661 gagtgcacgc ccccgcttgc ctccaggcac gtgctgaagc tgaccaacaa ctccaaccag  
721 ttccagaccg aggtcgggaa gcagctgatt tccggaaacc tggatgcacc cgaggggtggg  
781 ctggacgcca tgatgcaggt cggcgctgc ccggaggaaa tccgctggcg caacgtcacg  
30 841 cggctgctgg tgtttgccac tgatcacggc ttccatttcg cgggcgacgg aaagctgggc  
901 gccacagagt ccccaacga cggcgcgtgt cacctggagg acaacttgta caagaggagc  
961 aacgaattcg actaccatc ggtgggacag ctggcgacac agctggctga aaacaacatc  
1021 cagcccatct tcgctgtgac cagtaggatg gtgaagacct acgagaaact caccgagatc  
1081 atcccaagt cagcgtggg ggagctgtct gaggactcca gcaatgtggt ccatctcatt  
35 1141 aagaatgctt acaataaact ctctccagg gtcttcctgg atcacaacgc cctccccgac  
1201 accctgaaag tcacctacga ctcttctgac agcaatggag tgacgcacag gaaccagccc  
1261 agaggtgact gtgatggcgt gcagatcaat gtcccgatca ccttcagggt gaaggtcacg  
1321 gccacagagt gcatccagga gcagtcgttt gtcattccgg cgctgggctt caccgacata  
1381 gtgaccgtgc aggttcttcc ccagtgtagg tgccgggtgc gggaccagag cagagaccgc  
40 1441 agcctctgcc atggcaaggg cttcttgagg tgcggcatct gcaggtgtga cactggctac  
1501 attgggaaaa actgtgagtg ccagacacag ggccggagca gccaggagct ggaagggaagc  
1561 tgccggaagg acaacaactc catcatctgc tcagggctgg gggactgtgt ctgctggcag  
1621 tgctgtgccc acaccagcga cgtccccggc aagctgatat acgggcagta ctgctgagtg  
1681 gacaccatca actgtgagcg ctacaacggc caggtctgcg ggcggccggg gagggggctc  
45 1741 tgcttctgcy ggaagtgcg ctgccacccg ggctttgagg gctcagcgtg ccagtgcgag  
1801 aggaaccatc agggctgcct gaaccgcggc cgtgttgagt gtagtggctg tggcgggtgc  
1861 cgtgcaacg tatgcgagtg ccattcaggc taccagctgc ctctgtgcca ggagtgcccc  
1921 ggctgcccct caccctgtgg caagtacatc tctgctggc agtgctgaa gttcgaaaag  
1981 ggcccttttg ggaagaactg cagcgcggcg tgcggggcc tgcaagctgc gaacaacccc  
50 2041 gtgaagggca ggacctgcaa ggagagggac tcagagggct gctgggtggc ctacacgctg  
2101 gagcagcagg acgggatgga ccgctacctc atctatgtgg atgagagccg agagtgtgtg  
2161 gcaggcccca acatcgccgc catcgctggg gccaccgtgg caggcatcgt gctgatcggc  
2221 attctctctg tggctcatctg gaaggctctg atccacctga gcgacctccg ggagtacagg

2281 cgctttgaga aggagaagct caagtcaccag tggacaacatg ataatcccct tttcaagagc  
 2341 gccaccacga cggtcacgaa ccccaagttt gctgagagtt aggagcactt ggtgaagaca  
 2401 aggccgtcag gaccacccat gtctgcccc tccagcgcc gagacatggc ttggccacag  
 2461 ctcttgagga tgtcaccaat taaccagaaa tccagttatt ttccgccctc aaaatgacag  
 5 2521 ccatggccgg ccggtgcttc tgggggctcg tccggggggac agctccactc tgactggcac  
 2581 agtctttgca tggagacttg aggagggtt gaggttggtg aggttaggtg cgtgtttcct  
 2641 gtgcaagtca ggacatcagt ctgattaaag gtggtgccaa tttatttaca tttaaacttg  
 2701 tcagggtata aaatgacatc ccattaatta tattgttaat caatcacgtg tatagaaaaa  
 2761 aaaataaaac ttcaat

10 SEQ ID NO: 154

Amino acid sequence of human ITGB2 encoded by the DNA sequence shown in SEQ ID NO: 153.

15 MLGLRPPLLALVGLLSLGLCVLSQECTKFVSSCRECIESGPGCTWCQKLNFTGPGDPDSI  
 RCDTRPQLLMRGCAADDIMDPTSLAETQEDHNGGQKQLSPQKVTLYLRPGQAAAFNVTFR  
 RAKGYPIDLYLMDLSYSMLDDLNVKLLGGDLLRALNEITESGRIGFGSFVDKTVLPFV  
 NTHPDKLRNPKNKEKECQPPFAFRHVLKLTNNSNQFQTEVGKQLISGNLDAPEGGLDAM  
 MQVAACPEEIGWRNVTRLLVFATDDGFHAFAGDGKLGAILTPNDGRCHLEDNLYKRSNEFD  
 YPSVQQLAHKLAENNIQPIFAVTSRMVKTYEKLTEIPKSAVGELSEDSSNVVHLIKNAY  
 20 NKLSRVFLDHNLALPDTLKVTYDSFCSNGVTHRNQPRGDCDGVQINVPITFQVKVTATEC  
 IQEQSFVIRALGFTDIVTVQVLPQCECRCDQSRDRSLCHGKGFLECGICRCDTGYIGKN  
 CECQTQGRSSQBLEGSCRKDNNSIICSGLGDCVCGQCLCHTSDVPGKLIYQYCECDTIN  
 CERYNGQVCGGPGRGLCFCKCRCHPGFECSACQCERTTEGCLNPRVECSGRGRRCRCNV  
 CECHSGYQLPLCQECPCGSPSPCKYISCAECLKFEKGFPGKNCSAACPGQLSNNPVKGR  
 25 TCKERDSEGCWVAYTLEQQDGMRYLIYVDESRECVAGPNIAAIVGGTVAGIVLIGILL  
 VIWKALIHLSDLREYRRFEKEKLKSQWNNDNPLFKSATTTVMNPKFAES

SEQ ID NO: 155

gi|6680489|ref|NM\_008404.1| Mus musculus integrin beta 2 (Itgb2), mRNA

1 atgctgggccc tgcgcccctc actgctgctt gccctagctg gactgttctt cctgggatct  
 61 gctgtgtccc aggaatgcac caagtacaa gtcagcagtt gccgggactg tatccagtcg  
 121 gggcctgggt gttcctggtg ccagaagctg aacttcactg gaccaggaga acctgactcc  
 181 ttgcgctgtg acacacgggc acagctgctg ctgaagggtt gtccagccga tgatatcatg  
 241 gacccacagga gcatcgctaa tcctgagttc gaccaacggg ggcaacggaa acagctatct  
 301 ccacaaaaag tgacacttta ctgacgacca ggacaggctg ccgcatccta tgtgactttc  
 361 cggcggggcca agggataccc cattgatctg tactacctca tggatctctc ctactccatg  
 421 cttgatgacc tcaacaacgt caagaagctg ggcggggact tgctgcaggc cctcaacgag  
 481 atcacccagt ctggccgcat cggctttggg tcgtttgtgg acaagacggg gctgcctttt  
 541 gttaacaccc atcctgagaa gctgaggaa ccatgtccca acaaggagaa ggctgcccag  
 601 cccccatttg cctttcggca cgtgctcaag ttaaccgaca actccaacca gtttcagaca  
 661 gaggtcggca agcaactgat ttccggaaac ctggacgccc ctgagggtgg gctggatgcc  
 721 ataatgcaag ttgctgcatg tccggaggaa attggctggc gcaatgtcac gaggtgctg  
 781 gtgtttgcca cagacgatgg cttccacttt gctggtgatg gcaactggg tgccatcctg  
 841 acccccaatg atggccgctg ccacctggag gataacatgt acaagaggag caatgagttc  
 901 gactacccat ccgtgggtca gctggcacac aaactttccg agagcaacat ccagcccatc  
 961 tttcggtgga caaagaagat ttgaaaaacg tatgagaaac tcacggagat catcccraag  
 1021 tcagcagtg gggaaactgtc tgaactgctc agcaacgtgg tgcagctcat caagaatgcc  
 1081 tactataaac tctcctctag agtcttctc gaccacagca ccctcccga caccctgaaa  
 1141 gtcacctatg actcctctc cagtaatgga gcatcgagta taggcaaatc ccgtggggac  
 1201 tgtgatggcg tacagatcaa caaccgggtc acctccagg taaaggatcat ggcttccgag  
 1261 tgtatccagg agcagtcctt tgtcatccgg gactgggtt tcacgggatac agtgaccgtg  
 1321 caggctccgtc cccagtgtga gtgtcactgc cgggaccaga gtcggggagca gactctctgt  
 1381 ggaggcaagg gactcatgga gtgtggtatc tgcagggtgt agtctggcta cattgggaaa  
 1441 aactgtgagt gccagactca ggtcgggagc agccaggagc tggagagaaa ctgtcgggaag  
 1501 gacaatagtt ccatcggtg ctcagggtt ggggactgca tctgtgggca gtgtgtatgc

```

1561 cataccagtg acgtccccaa caaagagatc tttgggcaat actgcgagtg tgacaatgtc
1621 aactgtgaga gatataacag ccaagtctgc ggtgggctcag atcgggggttc ctgcaactgt
1681 ggcaaatgta gttgcaagcc cggttacgag ggctcggcct gccagtgtea gaggtccacc
1741 acgggctgtc tgaatgcacg gctggttagag tgcagtgagg gtggccactg ccaatgcaac
5 1801 aggtgcatat gtgacgaagg ctaccagcca ccgatgtgtg aggattgtcc cagctgtggc
1861 tcgcaactgca gggacaacca cacctcttgt gccagtgacc tgaagtttga taagggccct
1921 tttgagaaga actgtagtgt tcagtgtgct ggtatgacgc tgcagactat ccctttgaag
1981 aaaaagccct gcaaggagaa ggactcggaa ggctgttgga taacttacac tttgcagcag
2041 aaggacggaa ggaacattta caacatccat gtggaggaca gtctagagtg tgtgaagggc
10 2101 cccaatgtgg ctgccatcgt agggggcacc gtggtaggtg tcgtactgat tgggtgcctc
2161 ctccctggta tctggaaggc cctgaccacc ctgactgacc tcaggagagta caggcgcttt
2221 gagaaggaga aactcaagtc ccaatggaac aatgacaacc ccctcttcaa gagtgcagc
2281 acaacggtca tgaaccccaa gtttctgtaa agctagagca tgagttatca taatcaagca
2341 gatgtgaccc cctcagacca cgctcctccc cctctgcaaa cacaacgtgg cttacagctc
15 2401 accccagtgc tgccaaggat ccaaaagcct gctcggtttc ttncgcat tatatcaag

```

## SEQ ID NO: 156

Amino acid sequence of mouse ITGB2 encoded by the DNA sequence shown in SEQ ID NO: 155.

```

20 MLGLRPSLLLALAGLFFLGSAVSQECTKYKVSSCRDCIQSGPGCSWCQKLNFTGPGEPDS
LRCDTRAQLLLKGCPADDIMDPRS IANPEFDQRGQRKQLSPQKVTLYLPGQAAAFNVTF
RRAKGYPIDLYYLMDSYSMLDDLNNVKKLGGDLLQALNEITESGRIGFGSFVDKTVLPF
VNTHPKELRNPCPNKEKACQPPFAPRHVLKLTDSNQFQTEVGKQLISGNLDAPEGGLDA
IMQVAACPEEIGWRNVTRLVLFATDDGFHFAGDGKLGAILTPNDGRCHLEDNMYKRSNEF
25 DYPSVGQLAHKLSNESNIQPIFAVTKMKVKTYEKLTEIIPKSAVGELSDDSSNVVQLIKNA
YYKLSSRVFLDHSTLPD TLKVTYDSFCSNGASSIGKSRGDCDGVQINNPTVFQVKVMASE
CIQEQS FVIRALGFTDVTVTQVRPQCECHCRDQSRQSLCGGKGVMECGICRCESGYIGK
NCBCQTQGRSSQELERNCRKDNSSIVCSGLGDCICGQCVCHTSDVPNKEIFGQYCECDNV
NCERYNSQVCGGSDRGSCNCGKCSCKPGYEGSACQCQRSTTGCLNARLVECSGRGHCQCQCN
30 RCICDEGYQPPMEDCPSCGSHCRDNHTSCAECLKFDKGPFEKNCSVQCAGMTLQTIPLK
KKPCKEKDSEGWCWITYTLQKQDGRNIYNIHVEDSLECVKGPVAAIVGGTVVGVVLIGVL
LLVIWKALTHLTDLREYRRFEKEKLKSQWNNDNPLFKSATTTVMNPKFAES

```

## SEQ ID NO: 157

gi|34852368|ref|XM\_228072.2| Rattus norvegicus similar to Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/P150,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit) (LOC309684), mRNA

```

1 atggatggct ttggttgta tgtaacagga gccctgtccg aggaatgcac caagtacaaa
61 gtcagcaact gccgggactg tatccagtcg gggcctggct gctcgtggtg ccagaagctg
121 aacttcaccg gaccagggga gcctgactcc ttgcgctgtg acacgcgggc acagctgctc
40 181 ctcaaggggt gccagccga tgatataatg gaccccaaga gctttgctga tctccacccc
241 caatatcagg tgcaacggag tcaactgtct ccacaaaag tgacccttaa cttgcgacca
301 ggcagggctg ctgcattcaa tgtgactttc cgacgggcca agggctaccc cattgatctg
361 tactacctca tggacctctc ctactctatg ctgatgacc tgaacaatgt caagaagttg
421 ggtggtgatt tgcgtcaggc cctcaacgag atcacagagt ccggccgcat cggcttcggg
45 481 tccttcgtgg acaagacggg gctgcctttt gtcaacaccc atcccgagaa gctgaggaa
541 ccatgcccc acaaggagaa agcctgccag cctccgtttg cctttcgcca cgtgctcaag
601 ctaaaccgaca actccaacca gtttcagaca gaggtcggca agcaactgat ttccggaaac
661 ctggacgccc ctgagggcgg gctggatgcc ataatgcaag tggtgcctg tccgttctcc
721 tctgagccag gctgccagc agtggatctt ctggttccta aggcctttc cctctcagat
50 781 caagagcacg acggcttcca ctttgccggg gatgggaaac tgggtgccat cctgaccccc
841 aacgatggcc gctgccacct ggaggataac atgtacaaga ggagcaatga gttcgactac
901 ccgtcagtg gccagctggc ccacaaactt tccgagagca acatccagcc catctttgca

```

```

    961 gtgacaaaga agatgggtgaa aacctatgag aaactgacag agattatccc caagtcagcg
  1021 gtgggcgagc tgtctgacga ttccagtaac gtagtccagc ttatcaagaa agcctactac
  1081 aaactctcct ctagagtctt cctggaccac accaccatcc cggacaccct gaaagtcacc
    5 1141 tatgactcct tctgtaataa cagagtatcg agtataggca aatcccgagg ggactgtgac
    1201 ggtgtgcaga tcaacaaccc ggtcaccttc caggtaaagg tcacggcttc ggagtgtatc
    1261 caggagcagt cctttgtcat ccgggcgctg ggcttcaccg acacagtgac ggtacaggtc
    1321 catccccagt gcgagtgccg gtgcccggac cagagtcgga tgaggaatct ctgtggaggc
    1381 aagggagtca tggagtgtgg catctgcagg tgtgagtctg gctacattgg gaaaaactgt
    1441 gagtgtcaga cgcagggccg gagcagccag gagctggagg ggaactgccg gaaggacaat
  10 1501 agttccattg tgtgtctcggg gctgggggac tgcattctcg ggcagtgcgt gtgccacacg
    1561 agtgacatcc ccaacaaagt gatctttggg caatactgcg agtgtgacaa cttcaactgt
    1621 gagagatatg atggccaagt ctgggtgtgc ctaaagagag gctcctgctc ctgtggccag
    1681 tgtaattgca aggagggttt cgagggttct gcttgccagt gtcagaggtc taccacgggc
    1741 tgtctgaacg cacggctggt ggagtgcagt ggccgtggcc ggtgccaatg caacagatgc
  15 1801 atctgtgaga aaggttacca gccacctctg tgtgaagagt gtcccggctg ccccttgccc
    1861 tgcagacact acgtcttctg tgccgagtgc ctgaaatttg ataagggccc ctttcagaag
    1921 aattgtagtg ttcagtgtgc caatgtgacg ctgcagactg tccctttcaa gaaaaagccc
    1981 tgcaaggaga gggactcgga gggctgctgg ataacctaca ctttgacgca gaaggacgga
  20 2041 aacgcttaca acatccatgt ggacgagatg cgagagtgtg tgaaaggccc caacgtggct
    2101 gccatcatag ggggactgtg ggccggcgtt gtactgattg gtgtcctcct cctggctatc
    2161 tggaaggctc ttaccacact gactgacctc aatgaatata gacgctttga gaaggagaaa
    2221 ctcaagtccc agtgaacaa cgacaacccc ctcttcaaga gcgccacgac aacggtcatg
    2281 aaccctaagt ttgctgagag ctagagaagg agtcagagga gacccctcca gaccatgcct
    2341 cctccccctc gcaaatagaa tgtagcttac agctagcccc agtgcgtcca aggatccaaa
  25 2401 agcctacttt gtttctttcc gccattatat caaggctgcc aggggtttcca cagactcatc
    2461 ttccgacctt tacaatcttg ccacagagcc tgcagattgt tccggagtcc caacaggttc
    2521 cacacacggt ttctgtcata aagggaagac aggggtctca gtaaagggtg cccagctta
    2581 tttatattta aacttgtag cgtataaaac tactattata ttgttaacat cctgtccgtt
    2641 gtattatatg tgagtgtaaa actatatccc acatatatca gaatcatgtg tgtaaaaaata
  30 2701 ataaagcttc cattcagggc tgcagagatg gctcagtggg taagagcact gactgtctct
    2761 ccagaggtcc tgagttcaat tcccagcatc cacatgggtg ctcaacaaca tctgtaatgg
    2821 gatctgatgc ccttgtctgg tgtggctgaa gatagcaaca gtgtactcac atacataaaa
    2881 taaataaagc cttttaataa aaaaattaat aaagcttcca

```

SEQ ID NO: 158

- 35 Amino acid sequence of rat ITGB2 encoded by the DNA sequence shown in SEQ ID NO: 157.

```

  MDGFGCHVTGALSEECTKYKVSNCRDCIQSGPGCSWCQKLNFTGPGEPSLRCDTRAQLL
  LKGC PADDIMDPKSFADLHPQYQVQRSQLSPOKVTNLNLRPGQAAAFNVTFRRAKGYPIDL
  YYLMDLSYSMLDDLNNVKILGGDLLQALNEITESGRIGFGSFVDKTVLPFVNTHPEKLRN
  40 PCPNKEKACQPPFAFRHVLKLDNSNQFQTEVGKQLISGNLDAPEGGLDAIMQVAACPFS
  CEPGCPAVDLLVPKAFSLSDQEHDFHFAGDGKLGAILTPNDGRCHLEDNMYKRSNEFDY
  PSVGQLAHLKLSSENIQPIFAVTKKMVKTYEKLTEIIPKSAVGELSDDSSNVVQLIKAYY
  KLSSRVFLDHTTIPDTLKVITYDSFCNNRVSSIGKSRGDCDGVQINNPTVFQVKVTASECI
  QEQSFVIRALGFTDVTVTQVHPQCECQRDQSRMRNLGGKGVMCEGICRCESGYIGKNC
  45 ECQTQGRSSQLEGNCRKDNSSIVCSGLGDCICGQCVHTSDIPNKVIFGQYCECDNFNC
  ERYDGQVCGGLKRGSCSCGQCNCKEGFEGSACQCQRSTTGCLNARLVECSGRGRQCQCNRC
  ICEKGYPPLCEBCPGCPLPCSTYVFCAECLKFDKGPFPQKNCSVQCANVTLQTVPFKKKP
  CKERDSEGCWITYTLQKDGNAYNIHVDDRECVKGPNVAAIIGGTVAGVVLIGVLLLV
  WKALTHLTDLNEYRRFEKEKLKSQWNNDNPLFKSATTTVMNPKFAES

```

- 50 SEQ ID NO: 159

gi|20127646|ref|NM\_030569.2| Homo sapiens inter-alpha (globulin) inhibitor H5 (ITI5), mRNA



```

1  gtgtcccgcc  ggggtcccgga  gcgtcccgcg  ccctcgcccc  gccatgctcc  tgctgctggg
61  gctgtgacctg  ggggtgtctcc  tgtgtgtggg  gtcgcaggaa  gaggcgcaga  gctggggcca
121  ctcttcggag  caggatggac  tcagggtccc  gaggaagtc  agactgttgc  agaggctgaa
181  aaccaaacc  ttgatgacag  aattctcagt  gaagtctacc  atcatttccc  gttatgctt
5  241  cactacgggt  tcctgcagaa  tgctgaacag  agcttctgaa  gaccaggaca  ttgagttcca
301  gatgcagatt  ccagctgcag  ctttcatcac  caacttccact  atgcttattg  gagacaagggt
361  gtatcagggc  gaaattacag  agagagaaaa  gaagagtggg  gatagggtaa  aagagaaaag
421  gaataaaacc  acagaagaaa  atggagagaa  ggggactgaa  atattcagag  cttctgcagt
481  gattcccagc  aaggacaaa  ccgccttttt  cctgagttat  gaggagcttc  tgcagaggcg
10  541  cctgggcaag  tacgagcaca  gcacagcgt  gcggccccag  cagctgtccg  ggaggctgag
601  cgtgggcgtg  aatatcctgg  agagcgcggg  catcgcatcc  ctggagggtg  tgccgcttca
661  caacagcagg  cagaggggca  gtgggcgagg  ggaagatgat  tctgggcctc  ccccatctac
721  tgtcattaac  caaaatgaaa  catttgccaa  cataattttt  aaacctactg  tagtacaaca
781  agccaggatt  gcccagaatg  gaattttggg  agactttatc  attagatatg  acgtcaatag
15  841  agaacagagc  attggggaca  tccaggttct  aaatggctat  tttgtgcact  actttgtccc
901  taaagacctt  cctcctttac  ccaagaatgt  ggtattcgtg  cttgacagca  gtgcttctat
961  ggtgggaacc  aaactccggc  agaccaagga  tgccctcttc  acaattctcc  atgacctccg
1021  accccaggag  cgtttcagta  tcattggatt  tcccaaccgg  atcaaagtat  ggaaggacca
1081  cttgatatac  gtcactccag  acagcatcag  ggatgggaaa  gtgtacattc  accatatgtc
20  1141  acccactgga  ggcacagaca  tcaacggggc  cctgcagagg  gccatcaggc  tcctcaacaa
1201  gtacgtggcc  cacagtggca  ttggagaccg  gagcgtgtcc  ctcatcgtct  tcctgacgga
1261  tgggaagccc  acggtcgggg  agacgcacac  cctcaagatc  ctcaacaaca  ccgagagggc
1321  cgcccagggc  caagtctgca  tcttcaccat  tggcatcggc  aacgacgtgg  acttcaggct
1381  gctggagaaa  ctgtcgctgg  agaactgtgg  cctcacacgg  cgcgtgcacg  aggaggagga
25  1441  cgcaggctcg  cagctcatcg  ggttctacga  tgaaatcagg  accccgctcc  tctctgacat
1501  cgcgatcgat  tatccccca  gctcagtggt  gcaggccacc  aagacctgt  tccccacta
1561  cttcaacggc  tcggagatca  tcattgcggg  gaagctgggt  gacaggaagc  tggatcacct
1621  gcacgtggag  gtcaccgcca  gcaacagtaa  gaaattcctc  atcctgaaga  cagatgtgcc
1681  tgtgcccgt  cagaaggcag  ggaaagatgt  cacaggaagc  cccaggcctg  gaggcgatgg
30  1741  agagggggac  cccaaccaca  tcgagcgtct  ctggagctac  ctcaccacaa  aggagctgct
1801  gagctcctgg  ctgcaaagt  acgatgaacc  ggagaaggag  cggctgcggc  agcggggcca
1861  ggccctggct  gtgagctacc  gcttctcacc  tcccttcacc  tccatgaagc  tgagggggcc
1921  ggtcccacgc  atggacggcc  tggaggaggc  ccacggcatg  tcggctgcca  tgggacccga
1981  accggtgggt  cagagcgtgc  gaggagctgg  cagcgagcca  ggacctttgc  tcaagaagcc
35  2041  ataccagcca  agaattaaaa  tctctaaaa  atcagtggat  ggtgatcccc  actttgttgt
2101  ggatttcccc  ctgagcagac  tcaccgtgtg  cttcaacatt  gatgggcagc  ccggggacat
2161  cctcaggctg  gtctctgatc  acagggactc  tgggtgtcaca  gtgaacggag  agttaattgg
2221  ggcaccgcgc  cctccaaatg  gccacaagaa  acagcgcact  tacttgcgca  ctatcaccat
2281  cctcatcaac  aagccagaga  gatcttatct  cgagatcaca  ccgagcagag  tcactttgga
40  2341  tgggtggggac  agactgtgtc  tcccctgcaa  ccagagtgtg  gtgggtggga  gctgggggct
2401  ggaggtgtcc  gtgtctgcca  acgccaatgt  caccgtcacc  atccagggct  ccatagcctt
2461  tgtcatcctc  atccacctct  acaaaaagcc  ggcgccttc  cagcgacacc  acctgggttt
2521  ctacattgcc  aacagcgagg  gcctttccag  caactgccac  ggactgctgg  gtcagttcct
2581  gaatcaggat  gccagactca  cagaagaccc  tgcaggggcc  agccagaacc  tcactcacc
45  2641  tctgtctcct  caggtgggag  aggggcctga  ggccgtccta  acagtgaag  gccaccaagt
2701  cccagtggtc  tggaagcaaa  ggaagattta  caacggggaa  gagcagatag  actgctgggt
2761  tcccaggaac  aatgccgcca  aactgattga  cggggagtac  aaggattacc  tggcatccca
2821  tccatttgac  acagggatga  cacttgccca  gggaatgtcc  agggagctct  gaagctggca
2881  gccttaagaa  tgcaagtga  tgaaggacag  tgatgtgggg  aggccgtggg  gcagctcttt
50  2941  tcatggcttg  tacacgcctc  agctcctggc  aattagctgg  actccatgac  ccaccctgg
3001  tgcagcatag  atccgacgtc  tgtctggggc  aagggtaggg  gtgggtaggg  gcgggaagcc
3061  tgagtgcaaa  tgtcatttcc  ctctactgcc  tcttctgcc  tctccccacc  ctgcccacat
3121  ccacagaggg  gagagaaggg  tcatagctaa  atgcaacaaa  gtctgtatct  tgtcccaacc
3181  tgcttttctg  ttctgttagc  atatcataaa  gtaagccttt  ctggtg

```

55 SEQ ID NO: 160

Amino acid sequence of human ITIH5 encoded by the DNA sequence shown in SEQ ID NO: 159.

MLLLLGLCLGLSLCVGSQEAAQSWGHSSEQDGLRVPRQVRLQLRLKTKPLMTEFSVKSTI  
 ISRYAFTTVSCRMLNRASEDDIEFQMQUIPAAAFITNFTMLIGDKVYQGEITEREKKSGD  
 RVKEKRNKTTEENGEKGTEIFRASAVIPSKDKAAFFLSYEELLQRRLGKYEHSISVRPQQ  
 LSGRLSVGVNILESAGIASLEVLPLHNSRQRGSGRGEDDSGPPPSTVINQNETFANIIFK  
 5 PTVVQQARIAQNGILGDFIIRYDVNREQSIGDIQVLNGYFVHYFAPKDLPLPKNVVFL  
 DSSASMVGTKLRQTKDALFTILHDLRPQDRFSIIGFPNRIKVKDHLISVTPDSIRDGKV  
 YIHMSPTGGTDINGALQRAIRLLNKYVAHSGIGDRSVSLIVFLTDGKPTVGETHTLKIL  
 NNTREAARGQVCIFTIGIGNDVDFRLLEKLSLENCGLTRRVHEEDAGSQLIGFYDEIRT  
 PLLSDIRIDYPPSSVVQATKTLFPNYFNGSEII IAGKLVDRLDHLHVEVTASNSKKFII  
 10 LKTDVVPVRPQKAGKDVGTGSPRPGDGEGDPNHIERLWSYLTTKELLSSWLQSDDEPEKER  
 LRQRAQALAVSYRFLTPFTSMKLRGPVPRMDGLEEAHGMASAMGPEPVVQSVRGAGTQPG  
 PLLKKPYQPRIKISKTSVDGDPHFVDFPLSRLTVCFNIDGQPGDILRLVSDHRDSGVT  
 NGELIGAPAPPNGHKKQRTYLRTITILINKPERSYLEITPSRVILDGGDRLVLPQNQSVV  
 VGSWGLEVSVSANANVTVTIQSGIAFVILIHLYKKPAPFQRHHLGFYIANSEGLSSNCHG  
 15 LLGQFLNQDARLTEDPAGPSQNLTHPLLQVGEQPEAVLTVKGHPVPPVWKQRKIYNTEE  
 QIDCWFARNNAKLIDGEYKDYLAHPFDTGMTLGQMSREL

SEQ ID NO: 161

Amino acid sequence of human ITIH5, a soluble active secreted form derived from SEQ ID NO:160.

20 QSWGHSSEQDGLRVPRQVRLQLRLKTKPLMTEFSVKSTIISRYAFTTVSCRMLNRASEDQ  
 DIEFQMQUIPAAAFITNFTMLIGDKVYQGEITEREKKSGDRVKEKRNKTTEENGEKGTEIF  
 RASAVIPSKDKAAFFLSYEELLQRRLGKYEHSISVRPQQLSGRLSVGVNILESAGIASLE  
 VLPLHNSRQRGSGRGEDDSGPPPSTVINQNETFANIIFKPTVVQQARIAQNGILGDFIIR  
 YDVNREQSIGDIQVLNGYFVHYFAPKDLPLPKNVVFLDSSASMVGTKLRQTKDALFTI  
 25 LHDLRPQDRFSIIGFPNRIKVKDHLISVTPDSIRDGKVYIHMSPTGGTDINGALQRAI  
 RLLNKYVAHSGIGDRSVSLIVFLTDGKPTVGETHTLKILNNTREAARGQVCIFTIGIGND  
 VDFRLLEKLSLENCGLTRRVHEEDAGSQLIGFYDEIRTPLLSDIRIDYPPSSVVQATKT  
 LFPNYFNGSEII IAGKLVDRLDHLHVEVTASNSKKFII LKTDVVPVRPQKAGKDVGTGSPR  
 PGDGEGEDPNHIERLWSYLTTKELLSSWLQSDDEPEKERLRQRAQALAVSYRFLTPFTSM  
 30 KLRGPVPRMDGLEEAHGMASAMGPEPVVQSVRGAGTQPGPLLKKPYQPRIKISKTSVDGD  
 PHFVDFPLSRLTVCFNIDGQPGDILRLVSDHRDSGVTVNGELIGAPAPPNGHKKQRTYL  
 RTITILINKPERSYLEITPSRVILDGGDRLVLPQNQSVVGSWGLEVSVSANANVTVTIQ  
 GSIAPVILIHLYKKPAPFQRHHLGFYIANSEGLSSNCHGLLGQFLNQDARLTEDPAGPSQ  
 35 NLTHPLLQVGEQPEAVLTVKGHPVPPVWKQRKIYNTEEQIDCWFARNNAKLIDGEYKD  
 YLAHPFDTGMTLGQMSREL

SEQ ID NO: 162

gi|27369643|ref|NM\_172471.1| Mus musculus inter-alpha (globulin) inhibitor H5 (Itih5), mRNA

1 gccctgaaaa accttcgctg tcccagagag gtccgcgcgt ctggcttcgc catgctcctg  
 40 61 ctgctagggc tgtgcttggg gttgccccctc ttctcagagt cccaggaaga ggcaaggagt  
 121 tgggacgaca cctcggagca agtcgtgctc aggggtcccca ggcagctcag gctggtgcaa  
 181 agactgaaaa ccaagccctt gatggcagag ttttcggtga agtctaccat catttccgcg  
 241 tatgccttca ccacggtgtc ctgcaggatg ctcaacagag cttctgaaga ccaggaggct  
 301 gagttccaga tgcagattcc agaatacagct ttcataccca acttcacat gcttatagga  
 45 361 gacagcgtgt atcggggtga aattacacag aaagataaga aaagcagtga gagcgtaaaa  
 421 gataaaagga acagaacctc agacgataat gaagaaaacg ggagtgcacat gttcaaagcg  
 481 tcttttagtga ttcccagcaa ggacaaagct gctttcttcc tcagttatga agagcttctg  
 541 cagaggaggc tggggaaata cgagcatagc atcagcgtgc gccccagca gctggtgggg  
 601 aggcgtgact tggaggtgga catcctggag cgatcgggca tcacatcctt ggaagtgtg  
 50 661 cctctccaca acagcaggaa gaagggtagt gggaaaggcag aaggtgatgt gggctctccc  
 721 ccttctactc tcatcaacca aaatgagaca tttgccaaag tcattcttaa gcctactgtg  
 781 gtccaacaag ctaagattgc ccagaatggg atcttgggag atttcatcgt tcggtatgat

841 gtcgagagag agcagaacat tggtagacatc cagggttctga atggttattt cgtgcactac  
 901 tttgtctccta aaaacctccc tcctctaccc aagaacgtgg tcttcgtgct tgacatcagc  
 961 gcttccatgg tgggagcgaa actccagcag accagggagg ctctcgtcac aattctcaat  
 1021 gacctccgac ccagggatcg cttcaatatac attgggttctt ccaatcggat taaaatgtgg  
 5 1081 aaggaccact tactaccagt gactcctgac aacattagga atggcaagat ctacatgtac  
 1141 catctgtcac ccactggagg cacagacatc aatggggccc ttcaggctgc catcaaaactg  
 1201 ctcaacaact acgtggccca gaatgacatt gaagaccgaa gtgtgtccct catcatcttc  
 1261 ctaccgatg ggaaccacac ctttggggag accaataccc tcaagattct cagcaacacc  
 1321 aaggaagcca ccagaggtca gatctgcatc ttcaccgttg gcattggcga tgacgtggac  
 10 1381 ttcaaaactgt tagagaaact ctcgctggag aactgcggtc tcacgaggcg tgttcatgag  
 1441 gaggacaagg cgggcgacac gctcatcggg ttctatgatg agatccggac cccgcttctt  
 1501 tctgacatcc gcatagatta tcctcctgac gtagtagagc atgccaccaa gcacctgttc  
 1561 cctaactact tcaatggctc tgagatagtg attgcaggga agatgggtga caagaagttt  
 1621 gatcaactgc atgtggaggt cactgccagc aacagtaaga aatttgtcat cctaaaaaga  
 15 1681 gacatccccc tggagttccg gaagatgggg aatgatgtct cagtcacacc tggctcagcc  
 1741 agagatggcg ggaaggacct aaaccacatt gagcgacttt ggagctacct cactgtgaag  
 1801 gagctgctaa gctcctggag gcagagcaac agtgagcaag agaaagagca gctgaggcag  
 1861 aaggccagg acttagcctt gaactatcat ttcctcacc ccttcacctc catgaagctg  
 1921 aggaagccag ggctccgcac aaaccagctg gaggacacct atggcatgtc tgcaagctaca  
 20 1981 ggacctgcga ccgtgggtgca gaaccttcga gaggccggca agcagccaga acctgatctc  
 2041 aaaaaaacat atgacccaag aattaaaaac tctaaaacat cagtggatgg tgatcctcat  
 2101 tttgtagtgg atttccccct aagcaaacct actgtctgct ttaacatcga cggagagccg  
 2161 ggggacatcc tacgtctggt ctctgatcat ctgaactctg gtgtgactgt gaattggcgag  
 2221 cttattgggg cccccgcacc tccgaatggt cacaagaaac agcgcacctt ctccgcacc  
 25 2281 atcaccatcc tcatcaaccg gccagagaga tcttacctgg agatcacacc aagcagggtc  
 2341 atcctgggacg gtggggacac gctggtagtc cctgcaacc agagcgtggg agtagggagt  
 2401 cgaggattgg aggtgtcagt gtctgccaat gccaacatca ccgtgggtcat ccaggggcaac  
 2461 attgcctttg tcatcctcat ccacctgtac aaaaaccag cacccttcca gagagaccac  
 2521 ctgggcttct acatcgccaa cagcagaggt ctctccgaca actgccacgg actgctaggt  
 30 2581 cagttcctga accaggatgc caaacttgta ggagctcctg aggaatacgg caagaatctt  
 2641 agtaaccagc catttccctc ggacagaagg atgcctgagg ctatcctgaa ggtgaaaggg  
 2701 cgccgagttc cagtggctc gaaacaaaag aagatttaca acgggcaagc acaggtagac  
 2761 tcttggtttg acagaaacaa tgctgcaag ctgattgacg gtgtctataa ggactaccctg  
 2821 gcatctcatc cgtttgacac agagagtga ctgggcctga gcacgcccag gaaacctgag  
 35 2881 accgacaggc cccatgagga gagtgtctaa aagggaaagg atattgggtc tactctggac  
 2941 gcaactcttt ttatgtgtgg ctaccttggc tcttgctatg cgctgtactg ttagccccgc  
 3001 cccctgggtga gctggaggtc caatgtctct tagagcaatg agtcaggggca ggaagctgaa  
 3061 atgcaaatac ttcccttctt gccttctctca tetatactag ataaaaaaaa aaacaaacaa  
 3121 ac

40 SEQ ID NO: 163

Amino acid sequence of mouse ITIH5 encoded by the DNA sequence shown in SEQ ID NO: 162.

MLLLLGLCLGLPLFSESQEEARSWDDTSEQVLRVPRQLRLLQRLKTKPLMAEFSVKSTI  
 ISRYAFTTVSCRMNLRASEDQAEFQMIPESAFITNFTMLIGDSVYRGEITQDKKSSE  
 45 SVKDKRNRTSDDNBEENGSDMFKASLVIPSKDKAAFFLSYEELLQRLGKYEHSISVRPQQ  
 LVGRLTVEVDILERSGITSLEVLPLHNSRKKGSGKAEGDVGPPPSTLINQNETFAKVIFK  
 PTVVQQAQIAQNGILGDFIVRYDVEREQNIGDIQVLNGYFVHYFAPKNLPPLPKNVFVL  
 DISASMGAKLQQTREALVTILNDRPQDRFNIIGFSNRKMKWDLPLVTPDNIRNGKI  
 50 YMYHLSPTGGTDINGALQAAIKLLNNYVAQNDIEDRSVSLIIFLTDGKPTFGETNTLKIL  
 SNTKEATRQGQICIFTVGIGDDVDFKLLEKLSLENCGLTRRVHEEDKAGAQIGFYDEIRT  
 PLLSDIRIDYPPDVVEHATKTLFPNYFNGSEIIVAGKMVDKFPQLHVEVTASNSKKFVI  
 LKRDIPEVFRKMGNDVSVTPGSARDGGKDLNHIERLWSYLTVKELLSSWRQSNSEQEKEQ  
 LRQKAQDLALNYHFLTPFTSMKLRKPLRTNQLEDYTGMSAATGPATVVQNLRAGKQPE  
 PDLKKTYPRIKISKTSVDGDPHFVVDVDFPLSKLTVCFNIDGEPGDILRLVSDHLNSGVTV  
 55 NGELIGAPAPPNGHKKQRTYFRITITILINRPERSYLEITPSRVILDGGDRVLPCNQSVV  
 VGSRGLEVSVSANANITVVIQGNIAFVILIHLYKNPAPFQRDHLGFYIANSRGLSDNCHG  
 LLGQFLNQDAKLVGAPEEYGNLSNQFPFRAEGMPEAILKVGRVPVVKQRKIYNGQA

QVDCWFDRNNAAKLIDGVYKDYLAHPFDTESALGLSTPRKPETDRPHEESV

SEQ ID NO: 164

gi|20127658|ref|NM\_033101.2| Homo sapiens lectin, galactoside-binding, soluble, 12 (galectin 12) (LGALS12), mRNA

```

5      1 acaaacacctc gttggcccca cagggagcca gcctctggct tctctctgca atggccatgt
      61 gctgcagacc cggagtgggt agttagtgg ttaatgccag tcttctctcc ctggacactg
     121 agttctgctg acagcccccg cccagccaga gctctgctgt ataccaccgg gagggggct
     181 ggtgtggagc ctggaggtcg cccgctgccc tcctagggct gctccagaca gcattaaaac
     241 gctgcaggtc gcaggtgaga ctaacagctg ggagagctgc tccaggcatt taggaccctg
10    301 actggggcag atgagtcagc ccagtggggg cagggctcct ggaacgagga tctacagttg
     361 gagttgcccc actgtcatgt cacctggaga aaaactggac ccaattcctg acagcttcat
     421 tctgcaacca ccagtcttcc acccggtggt tccttatgtc acgacgattt ttggaggcct
     481 gcatgcaggc aagatggtca tgcgtcaagg agtggctcct ctagatgcac acaggtttca
     541 ggtggacttc cagtgtggct gcagcctgtg tccccggcca gatatcgctt tccacttcaa
15    601 ccctcgcttc cataccacca agccccatgt catctgcaac accctgcatg gtggacgctg
     661 gcaaaggagg gcccggtggc cccacctggc cctgcgaaga ggctccagct tcctcatcct
     721 ctttctcttc gggaatgagg aagtgaaggt gagtgtgaat ggacagcact ttctccactt
     781 ccgctaccgg ctcccactgt ctcatgtgga cacgctgggt atatttggtg acatcctggg
     841 agaggctggt ggattcctga acatcaatcc atttgtggag ggcagcagag agtaccagc
20    901 tggacatcct ttcctgctga tgagccccag gctggagggt ccctgctcac atgctcttcc
     961 ccagggtctc tcgctggggc aggtcatcat agtacgggga ctggctcttg aagagccgaa
    1021 gcattttact gtgagcctga gggaccaggc tgccccatgt cctgtgacac tcagggcctc
    1081 cttcgcagac agaactctgg cctggatctc ccgctggggg cagaagaaac tgatctcagc
    1141 ccccttcttc ttttaccccc agagattctt tgaggtgctg ctctgttcc aggaggagg
25    1201 gctgaagctg gcgctcaatg ggcaggggct gggggccacc agcatgaacc agcaggccct
     1261 ggagcagctg cgggagctcc ggatcagtgg aagtgtccag ctctactgtg tccactcctg
     1321 aggatggttc caggggaaat accgccagaa aacaagaagg tcagccctact cccaggggcc
     1381 cactctcctc ccctcattaa accatccacc tgacaccagc acatcaggcc tggttcacct
     1441 ctggggtcac gagactgagt ctacaggagc tttgggcctg agggaaaggc caagagtgca
30    1501 aaggttcctc gaactctgca cttcctcca ccaggagcct gggatatggc tccatctgcc
     1561 ttcagggcct ggactgcact cacagaggca agtgtgtgag actaacaag atactccaaa
     1621 atacaatggc ttaaagaatg tggtcattta ttctttatta tttatttatt tgtggtcaaa
    1681 taaataaata aggtta

```

SEQ ID NO: 165

35 Amino acid sequence of human LGALS12 encoded by the DNA sequence shown in SEQ ID NO: 164.

```

MSQPSGGRAPGTRIYSWSCPTVMSPGEKLDPIPD SFILQPPVFHPVVPYVTTFGGLHAG
KMVMLQGVVPLDAHRFQVDFQCGCSLCPRPDIAFHFNPRFHTTKPHVICNTLHGGRWQRE
ARWPHLALRRGSSFLILFLFGNEEVKVS VNGQHFLHFRYRLPLSHVDTLGIFGDILVEAV
40  GFLNINPFVEGSREYPAGHPFLMSRLEVPCSHALPQGLSPQQVIIVRGLVLQEPKHFT
VSLRDQAAHAPVTLRASFADRTLAWISRWGQKKLISAPFLFPQRF FEVLLLFQEGGLKL
ALNGQGLGATSMNQALEQLRELRISSVQLYCVHS

```

SEQ ID NO: 166

45 gi|15010855|gb|AF244977.1|AF244977 Homo sapiens galectin-12 isoform d mRNA, complete cds

```

      1 cctcagcttc ccagagtgt gagattacag gcatgagcca cagagcccag cccaggtaac
     61 ttctacatgg ggaaagggtga gcctgggaag gccattgtta agtgcctctc caggggtcca
    121 tttcagagtt agctgtcagt tcctaacttg tccattcttc ttgcctccca gttcctagtc

```

```

181 tccagccact caacagtcct cagaaagccc cagtaccagg cttctgggga ccagggtgac
241 atctagtggg tggcaagcct ccctgggggc actgcagcct gttgtgtcct aggagctggg
301 gaaaggcaac tggctggtgc tgaatttggc aagggagcgg aatttgactt ctctggtgta
361 atgcagcttt gccgtgtgac ggtccaggaa aggggattag gtggaagaaa atatttcagt
5 421 gaacactgat ttttacctat aaggaatttt ctgtttggag aggaaaagt gaggtttatt
481 ctgctccttt cctccactc gcctgacata gaggtcccta agccctttct ccaaacctgc
541 atggatgagt ttcttttctt gttcaggtgg ttccttatgt caccagcatt tttggaggcc
601 tgcattgcagg caagatgggc atgctgcaag gagggtgccc tctagatgca cacaggtttc
661 agtgaggactt ccagtgtggc tgcagcctgt gtccccggcc agatatacgc ttcacttca
10 721 accctcgctt ccataccacc aagccccatg tcatctgcaa caccctgcat ggtggacgct
781 ggcaaaggga ggcccggtgg cccacactgg cctgcgaag aggcctccagc ttcctcctcc
841 tctttctctt cggaatgag gaagtgaagg tgagtgtgaa tggacagcac tttctccact
901 tccgtaccg gctccactg tctcatgtgg acacgctggg tatatttggg gacatcctgg
961 tagaggtgtg tggattcctg aacatcaatc catttgtgga gggcagcaga gaggaccag
15 1021 ctggacatga ggtgccctgc tcacatgctc tccccaggg tctctgcct gggcaggtca
1081 tcatagtacg gggactgggc ttgcaagagc cgaagcattt tactgtgagc ctgaggggacc
1141 aggctgcccc tgctcctgtg acactcaggg cctccttcgc agacagaact ctggcctgga
1201 tctcccgtg ggggcagaag aaactgatct cagccccctt cctcttttac cccagagat
1261 tctttctggt gctgctcctg ttccaggagg gagggtgaa gctggcgctc aatgggcagg
20 1321 ggctgggggc caccagcatg aaccagcagg ccctggagca gctgcgggag ctccgatca
1381 gtggaagtgt ccagctctac tgtgtccact cctgaggatg gttccagggg aaataccgcc
1441 agaaaacaag aaggtcagcc cactcccagg gccccactct cctccccctc ttaaaccatc
1501 cactgcacac cagcacatca ggcctgggtc acctctgggg tcacgagact gaggctacag
1561 gagctttggg cctgagggaa ggcacaagag tgcaaagggt cctcgaaact tgcaccttc
25 1621 tccaccagga gcctgggata tggtccatc tgcttcagg gcctggactg cactcacaga
1681 ggcaagtgtt gtagactaac aaagatactc caaaatacaa tggcttaaag aatgtggtca
1741 tttattcttt attatttatt tatttgtggt caaataaata aataaggtta

```

SEQ ID NO: 167

30 Amino acid sequence of human LGALS12 variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 166.

```

MVMLQGVVPLDAHRFQVDFQCGCSLCPRPDI AFHFNPRFHTTKPHVICNTLHGGRWQREA
RWPFLALRRGSSFLILFLGNEEVKVS VNGQHFLHFRYRLPLSHVDTLGI FGDILVEAVG
FLNINPFVEGSREYPAGHEVPCSHALPQGLSPGV IIVRGLVLQEPKHFTVSLRDQAAHA
PVTLRASFADRTLAWISRWGQKKLISAPFLFYPQRF FEVLLLFQEGGLKLALNGQGLGAT
35 SMNQQALEQLRELRI SGSVQLYCVHS

```

SEQ ID NO: 168

gi|15010853|gb|AF244976.1|AF244976 Homo sapiens galectin-12 isoform c mRNA, complete cds

```

1 acaaaccttc gttggcccca caggagacca gcctctggct tctctctgca atggccatgt
40 61 gctgcagacc cggagtgggt agttagtgtg ttaatgccag tcttctctcc ctggacactg
121 agttctgctg acagcccccg cccagccaga gctctgctgt ataccaccgg gaggggggct
181 ggtgtggagc ctggagggtcg cccgtgccc tcctagggct gctccagaca gcattaaaac
241 gctgcaggtc gcagggtgaga ctaacagctg ggagagctgc tccaggcatt taggacctg
301 actggggcag atgagtcagc ccagtggggg cagggtcctt ggaacgagga tctacagttg
45 361 gaggttgccc actgtcatgt cacctggaga aaaactggac ccaattcctg acagcttcat
421 tctgcaacca ccagtcttcc acccggtggt tccttatgtc acgacgattt ttggaggcct
481 gcatgcaggc aagatgggtc tgctgcaagg agtgggtcct ctagatgcac acaggtttca
541 ggtggacttc cagtgtggct gcagcctgtg tccccggcca gatatacgc ttcacttcaa
601 cctcgccttc cataccacca agccccatgt catctgcaac accctgcatg gtggacgctg
50 661 gcaaaggagg gcccgggtgg cccacctggc cctgcgaaga ggctccagct tcctcactct
721 ctttctcttc gggaatgagg aagtgaaggt gagtgtgaat ggacagcact tctcactt
781 ccgctaccgg ctcccactgt ctcattgtgga cagctgggt atatttgggt acatcctgg

```

```

      841 agaggctgtt ggattcctga acatcaatcc atttgtggag ggcagcagag agtaccacgc
      901 tggacatgag gtgccctgct cacatgctct tccccagggc ctctcgccctg ggcagggtcat
      961 catagtacgg ggactggctct tgcaagagcc gaagcatttt actgtgagcc tgagggacca
5    1021 ggctgcccat gctcctgtga cactcagggc ctcttcgca gacagaactc tggcctggat
      1081 ctcccgtgg gggcagaaga aactgatctc agcccccctc ctcttttacc cccagagatt
      1141 ctttgaggtg ctgctcctgt tccaggaggg agggctgaag ctggcgctca atgggcaggg
      1201 gctgggggcc accagcatga accagcaggg cctggagcag ctgcgggagc tccggatcag
      1261 tggaagtgtc cagctctact gtgtccactc ctgaggatgg tccagggga aataaccgcca
10   1321 gaaaacaaga aggtcagccc actcccaggg cccactctc ctcccctcat taaaccatcc
      1381 acctgacacc agcacatcag gcctggttca cctctggggc cagcagactg agtctacagg
      1441 agctttgggc ctgagggaag gcacaagagt gcaaagggtc ctggaactct gcaccttctt
      1501 ccaccaggag cctgggatat ggctccatct gccttcaggg cctggactgc actcacagag
      1561 ccaagtgttg tagactaaca aagatactcc aaaatacaat ggcttaaaga atgtgggtcat
      1621 ttattcttta ttatttattt atttgtggtc aaataaataa ataaggtta

```

## 15 SEQ ID NO: 169

Amino acid sequence of human LGALS12 variant ORF number 2 encoded by the DNA sequence shown in SEQ ID NO: 168.

```

20   MSQPSGGRAPGTRIYSWSCPTVMSPEKLDPIPDSEFILQPPVFHPVVPYVTTIFGGLHAG
      KMVMLQGVVPLDAHRFQVDFQCGCSLCPRPDIAFHFNPRFHTTKPHVICNTLHGGRWQRE
      ARWPHLALRRGSSFLILFLFGNEEVKVSUNGQHLHFRYRLPLSHVDLTGIFGDILVEAV
      GFLNINPFVEGSREYPAGHEVPCSHALPQGLSPGQVIIVRGLVLQEPKHFTVSLRDQAAH
      APVTLRASFADRTLAWISRWGQKKLISAPFLFPQRF FEVLLLFQEGGLKLALNGQGLGA
      TSMNQALEQLRELRISSGVQLYCVHS

```

## SEQ ID NO: 170

## 25 gi|15010851|gb|AF244975.1|AF244975 Homo sapiens galectin-12 isoform b mRNA, complete cds

```

      1 cctcagcttc ccagagtgt gagattacag gcatgagcca cagagcccag cccaggtaac
      61 ttctacatgg ggaaagggtga gcctgggaag gccattgtta agtgcctctc caggggtcca
      121 tttcagagtt agctgtcagt tcctaaacttg tccattcttc ttgcctccca gttcctagtc
30   181 tccagccact caacagtctc cagaaagccc cagtaccagg cttctgggga ccagggtgac
      241 atctagtggg tggcaagcct ccctgggggc actgcagcct gttgtgtcct aggagtggg
      301 gaaaggcaac tggctgggtgc tgaatttggc aaggagcggg aatttgactt ctctgggtga
      361 atgcagcttt gccgtgtgac ggtccaggaa aggggattag gtggaagaaa atatttcagt
      421 gaacactgat tttacctat aaggaaattt ctgtttggag aggaaaagtt gagttttatt
35   481 ctgctccttt cctcccactc gcctgacata gaggtcccta agccctttct ccaaacctgc
      541 atggatgagt ttcttttctt gttcaggtyg ttcttatgt cagcagatt ttggaggcc
      601 tgcattgcagg caagatgggc atgctgcaag gagtgggtccc tctagatgca cacaggttc
      661 aggtggactt ccagtgtggc tgcagcctgt gtccccggcc agatatcgcc ttcacttca
      721 accctcgctt ccataaccac aagccccatg tcatctgcaa caccctgcat ggtggacgct
40   781 ggcaaggga ggcccggtgg cccacctgg ccctgcgaag aggtccagc ttctcatcc
      841 tctttctctt cgggaatgag gaagtgaagg tgagtgtgaa tggacagcac tttctccact
      901 tccgctaccg gctcccactg tctcatgtgg acacgctggg tatatttggg gacatcctgg
      961 tagaggctgt tggattcctg aacatcaatc catttggtga gggcagcaga gattaccag
45   1021 ctgggacatc ttctctgctg atgagcccca ggctggagggt gccctgctca catgctcttc
      1081 cccagggtct ctgcctggg caggatcatc tagtacgggg actggtcttg caagagccga
      1141 agcattttac tgtgagcctg agggaccagg ctgcccctgc tcctgtgaca ctcagggcct
      1201 ccttcgcaga cagaactctg gcctggatct cccgctgggg gcagaagaaa ctgatctcag
      1261 ccccttctct cttttacccc cagagattct ttgagggtgt gtcctgttcc caggaggggag
      1321 ggctgaagct ggcgctcaat gggcaggggc tggggggccac cagcatgaac cagcaggccc
50   1381 tggagcagct gcgggagctc cggatcagtg gaagtgtcca gctctactgt gtccactcct
      1441 gaggatgggt ccaggggaaa taccgcagga aaacaagaag gtcagccccc tcccaggggc
      1501 ccactctctt cccctcatta aaccatccac ctgacaccag cacatcaggg ctggttcacc

```

1561 tctgggggtca cgagactgag tctacaggag ctttgggcct gagggaaggc acaagagtgc  
 1621 aaaggttccct cgaactctgc accttcctcc accaggagcc tgggatatgg ctccatctgc  
 1681 cttcagggcc tggactgcac tcacagaggc aagtgttgta gactaacaaa gatactccaa  
 1741 aatacaatgg cttaaagaat gtggtcattt attctttatt atttatttat ttgtggtcaa  
 5 1801 ataaataaat aagggtta

SEQ ID NO: 171

Amino acid sequence of human LGALS12 variant ORF number 3 encoded by the DNA sequence shown in SEQ ID NO: 170.

10 MVMLQGVVPLDAHRFQVDFQCGCSLCPRPDIAPHFNPRFHTTKPHVICNTLHGGRWQREA  
 RWPHLALRRGSSFLILFLFGNNEEVKVSUNGQHLHFRYRLPLSHVDTLGI FGDILVEAVG  
 FLNINPFVEGSREYPAGHPFLMSPRLEVPCSHALPQGLSPGQVIVRGLVLQEPKHFTV  
 SLRDQAAHAPVTLRASFAVRTLAWISRWGQKKLISAPFLFYPQRF FEVLLLFQEGGLKLA  
 LINGQLGATSMNQALEQLRELRI SGSVQLYCVHS

SEQ ID NO: 172

15 gi|1878242|gb|AF310686.1|AF310686 Homo sapiens galectin-12 splice form 1 mRNA,  
 complete cds, alternatively spliced

1 agttggagtt gcccactgt catgtcacct ggagaaaaac tggacccaat tcctgacage  
 61 ttcatctctgc aaccaccagt cttccaccg gtggttcctt atgtcacgac gatttttggga  
 121 ggctgcatg caggcaagat ggtcatgctg caaggagtgg tcctctctaga tgcacacagg  
 20 181 tttcaggtgg acttcagtg tggctgcagc ctgtgtcccc ggccagatat cgccttccac  
 241 ttcaacctc gcttcctac caccaagccc catgtcatct gcaacacctt gcatggtgga  
 301 cgctggcaaa gggaggccc gtggccccc ctggccctgc gaagaggctc cagcttctc  
 361 atcctcttctc tcttcgggaa tgaggaaagt aagggtgagt tgaatggaca gcactttctc  
 421 cacttccgct accggctccc actgtctcat gtggacacgc tgggtatatt tggtgacatc  
 25 481 ctggtagagg ctgttgatt cctgaacatc aatccatttg tggagggcag cagagagtac  
 541 ccagctggac atcctttctt gctgatgagc cccaggctgg aggtgccctg ctcacatgct  
 601 cttccccagg gtctctcgc tgggcaggtc atcatagtag ggggactggt cttgcaagag  
 661 ccgaagcatt ttactgtgag cctgagggac caggctgccc atgtctctgt gacactcagg  
 721 gcctccttcg cagacagaac cctggcctgg atctcccgtc gggggcagaa gaaactgatc  
 30 781 tcagccccc tctctttta ccccagaga ttctttgagg tgctgctcct gttccaggag  
 841 ggagggtgta agctggcgct caatgggcag gggctgggg ccaccagcat gaaccagcag  
 901 gccctggagc agctgcggga gctccggatc agtggaaagt tccagctcta ctgtgtccac  
 961 tcctgaggat ggttccagg aaataccgcc agaaaacaag aaggctagcc cactcccagg  
 1021 gcccactct cctcccctca ttaaaccatc cacctgacac cagcacatca ggctggttc  
 35 1081 acctctgggg tcacgagact gagtctacag gagctttggg cctgagggaa ggcacaagag  
 1141 tgcaaagggt cctcgaactc tgcaccttc tccaccagga gcctgggata tggctccatc  
 1201 tgccttcagg gcctggactg cactcacaga ggcaagtgtt gtagactaac aaagatactc  
 1261 caaaatacaa tggcttaaag aatgtggtca tttattcttt attatttatt tatttgggt  
 1321 caaataaata aataagggtta tttattt

40 SEQ ID NO: 173

Amino acid sequence of human LGALS12 variant ORF number 4 encoded by the DNA sequence shown in SEQ ID NO: 172.

45 MSPGEKLDPI PDSFILQPPVFHPVVPYVTTIFGGLHAGKMVMLQGVVPLDAHRFQVDFQC  
 GCSLCPRPDIAPHFNPRFHTTKPHVICNTLHGGRWQREARWPHLALRRGSSFLILFLFGN  
 EEVKVSUNGQHLHFRYRLPLSHVDTLGI FGDILVEAVGFLNINPFVEGSREYPAGHPFL  
 LMSRLEVPCSHALPQGLSPGQVIVRGLVLQEPKHFTVSLRDQAAHAPVTLRASFAVRT  
 LAWISRWGQKKLISAPFLFYPQRF FEVLLLFQEGGLKLA LINGQLGATSMNQALEQLRE  
 LRISGSVQLYCVHS

## SEQ ID NO: 174

gi|11878244|gb|AF310687.1|AF310687 Homo sapiens galectin-12 splice form 2 mRNA, complete cds, alternatively spliced

```

5      1 agttggagtt gccccactgt catgtcacct ggagaaaaac tggacccaat tcctgacagc
      61 ttcattctgc aaccaccagt cttccaccgg gtggttcctt atgtcacgac gattttttgga
      121 ggctgcatg caggcaagat ggtcatgctg caaggagtgg tccctctaga tgcacacagt
      181 aggtttcagg tggacttcca gtgtggctgc agcctgtgtc cccggccaga tatcgcttc
      241 cacttcaacc ctgcttcca taccaccaag ccccatgtca tctgcaacac cctgcatggt
      301 ggacgctggc aaaggaggc cgggtggccc cacctggccc tgcgaagagg ctccagcttc
10     361 ctcatcctct ttctcttcgg gaataggaa gtgaaggatga gtgtgaatgg acagcacttt
      421 ctccacttcc gctaccggct cccactgtct catgtggaca cgtgggtat atttggtagac
      481 atcctggtag aggtgtgtgg attcctgaac atcaatccat ttgtggaggg cagcagagag
      541 taccagctg gacatccttt cctgctgatg agccccaggc tggagggtgcc ctgctcacat
      601 gctcttcccc agggctctct gcctgggcag gtcacatag tacggggact ggtcttgcaa
15     661 gagccgaagc attttactgt gacgtgagg gaccaggctg cccatgctcc tgtgacactc
      721 agggctctcc tcgcagacag aactctggcc tggatctccc gctgggggca gaagaaactg
      781 atctcagccc ccttctctct ttacccccag agattctttg aggtgctgct cctgttccag
      841 gagggagggc tgaagctggc gctcaatggg caggggctgg gggccaccag catgaccag
      901 caggccctgg agcagctgag ggagctccgg atcagtggaa gtgtccagct ctactgtgtc
20     961 cactcctgag gatggttcca gggaaatacc gccagaaaac aagaaggtca gccactccc
      1021 agggcccccac tctcctcccc tcattaaacc atccacctga caccagcaca tcaggcctgg
      1081 ttcacctctg gggtcacgag actgagtcta caggagcttt gggcctgagg gaaggcacia
      1141 gagtgcaaag gttcctcgaa ctctgcacct tcctccacca ggagcctggg atatggctcc
      1201 atctgccttc agggcctgga ctgcactcac agaggcaagt gttgtagact aacaaagata
25     1261 ctccaaaata caatggctta aagaatgtgg tcatttatte tttattattt atttatttgt
      1321 ggtcaaataa ataaataagg ttatttatte

```

## SEQ ID NO: 175

Amino acid sequence of human LGALS12 variant ORF number 5 encoded by the DNA sequence shown in SEQ ID NO: 174.

```

30     MSPGEKLDPIPDSEFILQPPVFHPVVPYVTTIFGGLHAGKMVMLQGVVPLDAHSRFQVDFQ
      CGCSLCPRPDIAFHFNPRFHTTKPHVICNTLHGGRWQREARWPHLALRRGSSFLILFLFG
      NEEVKVSVNGQHFLHFRYRLPLSHVDTLGIFGDILVEAVGFLNINPFVEGSREYPAGHPF
      LLMSPLREVPFCSHALPQGLSPGQVIVRGLVLQEPKHFTVSLRDQAAHAPVTLRASFAADR
35     TLAWISRWGQKKLISAPFLFYPQRFEEVLLLFQEGGLKLALNGQGLGATSMNQQALEQLR
      ELRISGSVQLYCVHS

```

## SEQ ID NO: 176

gi|15010845|gb|AF244978.1|AF244978 Mus musculus galectin-12 mRNA, variant a, complete cds, alternatively spliced

```

40     1 cacgtgtacc accagcagca gcagcttccc ttccactat ccgctgaagg cccagctggt
      61 gcagaggctg ggaaaagtga gtccattccg tggggaatcg cgttgctgga atcaggatca
      121 gaggtgagt tgaagttgcc cctccccat gtcaactgac gaacacctgg acccgattcc
      181 tgacagcttc atcctgcagc cgccagctct ccaccgggtg attccttatg gcacaaccaat
      241 ttttgggtggc ctgtatgcag gcaagatggg cacgctgcag ggtgtgggtc ctctgcatgc
45     301 aaggagggtt cagggtggact tccagtgtgg gtgctgcctg catectcagc cagatgttgc
      361 cttccgcttc agccctcgat tctacactgt caagcccat gtcacttgca acaccacca
      421 aggtggactc tggcaaaaag agatacgggt gccaggggtc gccctgcaga gaggggatag
      481 cttcctcatt ctctttctct ttgagaacga agaggtgaag gtgagtgtaa atggccagca
      541 ctttcttcac taccgctacc ggctccact gtcacgggtg gatacccttg acatatctgg
      601 tgacatcttg gtaaaggctg ttggattcct gaacatcaat ccatttgtgg agggtagcag

```



```

661 agagtatcca gttggatata ccttcctgct gtatagcccc aggctggagg tgccttgcctc
721 acgtgctctt cctcgggggc tctggcctgg gcaagtcatt gtagttcgag gactgggtctt
781 gaaagagccg aaagatttta ccctgagcct gaaggatggg accacccatg ctctgtgtac
841 actcagggtc tccttcacag acagaacact ggcttgggtc tcctcctggg gacgaaagaa
5 901 gctgatctcc gcccccttcc tctttcaccc ccagcgattc ttcgaggtag tgcttctgtg
961 ccaagaggga gggctgaagt tggcactcaa tgggcagggg ctggggggcca ccagcctgga
1021 ccagaaagcc ctggagcagc tgcgggagct caggatcagt ggaaatgtcc acctctactg
1081 tgtccactgc taagaagggc tccaagggca ccccgcagc agaagaggag gggcggtctg
1141 aaccagggtc ccacagttac agcccattct cctattctca agggagaaac tgacctggca
10 1201 tcactttgtc aggccgagct cacatctcaa agggcgcgcg cacacacaca c

```

SEQ ID NO: 177

Amino acid sequence of mouse LGALS12 encoded by the DNA sequence shown in SEQ ID NO: 176.

```

15 MSTDEHLDPIPDSPFILQPPVFHPVPIPYGTTIFGGLYAGKMVTLQGVVPLHARRFQVDFQC
GCCLHPQPDVAFRFSRPFYTVKPHVICNTHQGGLWQKEIRWPGVALQRGDSFLILFLFEN
EEVKVSVNGQHFLHYRYRLPLSRVDLTLDISGDILVKAVGFLNINPFVEGSREYPVGYPFLL
LYSPRLLEVPCSRALPRGLWPGQVIVVRGLVLKEPKDFTLSLKDGTTTHAPVTLRASFTDRT
20 LAWVSSWGRKKLISAPFLFHPQRFVEVLLLCQEGGLKALNGQGLGATSLDQKALEQLRE
LRISGNVHLYCVHC

```

SEQ ID NO: 178

gi|34861855|ref|XM\_219545.2| Rattus norvegicus similar to galectin-12 (LOC293710), mRNA

```

25 1 atgtcaactg acgaacacct ggacccgatc cctgacagct tcatcctgca gccgccagtc
61 ttccaccctg tgggtgctga gcttgctgaa gggaaagaca tggctccagt tttcttatac
121 agtgatcaga atggaaaaac agccggggcag atcctctata ctggggccat ggaccgagca
181 agcattgtga ttcttattgt cacaacaatt ttcggtggcc tgtatgcagg caagatgatc
241 atgctgcagg gtgtggtccc tcgccatgca cggaggtttc aggtggactt ccagtgtggg
301 tgctgcctgc atcctcggcc agatgttgcc ttccacttca gccctcgctt ctacactgtc
30 361 aagcccccag tcacttgcaa caccctccaa ggtggactct ggcagaaaga ggtccggtgg
421 ccaggaatcg ccctgcagaa aggggctagc ttctcatcc tctttctctt tgacaatgaa
481 gaggtgaagg tgagtgtgaa cggacagcac tttcttccact accgctatcg gctcccactg
541 tcacgggtgg atactcttga tatatctggc gacatcttgg taaaggctgt tggattcctg
601 aacatcagtc cattcgtgga gggtagcaga gagtatccag ttggatatcc cctcctgctg
35 661 tatagcccca ggctgtacca agcatcagcc caggaaaacc atccagcccc agacctccag
721 ctgctcatcg gtggcagaat gacaaaccac tgtataacctg ctacatacag ttcagagcac
781 actctggaac aaggacctgt gcacagccca gagagcctga agaagctctc agcaggctta
841 catttcaccc tgagcctgag ggatggggcc acccatgttc ctgtgacact cagggttcc
901 ttcacagaca gaacactggc ctgggtctcc tcctggggac gaaagaagct gatctcagcc
40 961 cccttctctt tttatcccca gcgattcttc gaggttctgc ttttgtgcca agagggaggg
1021 ctgaagctgg cactcaatgg gcatgggctg gggggccacca gcctggatca gaaagccctg
1081 gagcagctgc gggacctcag gatcagtgga aggtccaca gttacagccc caccctgcct
1141 ttctcactgg agaacttgcc ctggcatcac tgtttcaggc ctagtccaca tctggtcact
1201 caactagatt gtggagcctt tgcccaaagt gaacaaatat cacagctgct ttcattggtt
45 1261 tgggtcaagg gaaccagcca catcccagaa aaacacactg ataatgaggc cacattctgg
1321 tcacctaaaca ataactctgc ataa

```

SEQ ID NO: 179

Amino acid sequence of rat LGALS12 encoded by the DNA sequence shown in SEQ ID NO: 178.

MSTDEHLDPIDPSFILQPPVFHPVGAELAEKGDMAPVFLYSDQNGKTAGQIILYTGAMDRA  
 SIVIPYVTTIFGGLYAGKMIMLQGVVPRHARRFQVDFQCGCCLHPRPDVAFHFSRPFYTV  
 KPHVICNTLQGGWLQKEVRWPGIALQKGASFLILFLFDNEEVKVSVNGQHFLHYRYRLPL  
 SRVDTLDISGDILVKAVGFLNISPFEVGSREYPVGYPPLLLYSPRLYQASGQENHPAPDLQ  
 5 LLIIGGRMTNHCIPATYSSEHTLEQGPVHSPESLKKLSAGLHFTLSLRDGATHVPVTLRAS  
 FTDRTLAWVSSWGRKKLISAPFLFYQRFEEVLLLCQEGGLKLALNGHGLGATSLDQKAL  
 EQLRDLRISGRLSYSPTLPFSLENLPWHHCFRPSSHLVTQLDCGAFAQSEQISQLLSWF  
 WSKGTSHIPEKHTDNEATFWSPNNLA

SEQ ID NO: 180

10 gi|24475648|ref|NM\_021077.2| Homo sapiens neuromedin B (NMB), mRNA

1 cgcgcgcccc aacgaagccg cggcccgggc acagccatgg cccggcgggc ggggggcgct  
 61 cggatgttcg gcagcctcct gctcttcgcc ctgctcgtcg cggcgctcgc cccgctcagc  
 121 tgggatctcc cggagccccg cagccgagcc agcaagatcc gagtgcactc gcgaggcaac  
 181 ctctgggcca cgggtcactt catgggcaag aagagtctgg agccttccag cccatcccat  
 15 241 tggggacagc tccccacacc tccctgagg gaccagcgac tgcagctgag tcatgatctg  
 301 ctcggaatcc tccctgctaaa gaaggtctcg ggcgtgagcc tcagccgccc cgcaccccaa  
 361 atccagtaca ggaggctgct ggtacaaata ctgcagaaat gacaccaata ataggggagc  
 421 acacaacagc gtggcttaga ttgtgcccac ccagggaagg tgctgaatgg gacctgttg  
 481 atggcccat ctggatgtaa atcctgagct caaatctctg ttactccatt actgtgattt  
 20 541 ctggctgggt caccagaaat atcgctgatg cagacacaga ttatgttccct gctgtatttc  
 601 ctgcttcctt gttgaattgg tgaataaaac cttgctcttt

SEQ ID NO: 181

Amino acid sequence of human NMB encoded by the DNA sequence shown in SEQ ID NO: 180.

25 MARRAGGARMFGSLLLFALLAAGVAPLSWDLPEPRSRASKIRVHSRGNLWATGHFMGKKS  
 LEPSSPSHWGQLPTPLRDQRLQLSHDLLGILLKKALGVSLSRPAPQIQYRLLLVQILQ  
 K

SEQ ID NO: 182

30 Amino acid sequence of human NMB, a soluble active secreted form derived from SEQ ID NO: 181.

APLSWDLPEPRSRASKIRVHSRGNLWATGHFMGKKSLEPSSPSHWGQLPTPLRDQRLQL  
 SHDLLGILLKKALGVSLSRPAPQIQYRLLLVQILQK

SEQ ID NO: 183

gi|13386017|ref|NM\_026523.1| Mus musculus neuromedin B (Nmb), mRNA

35 1 gatttggcgc gcttcgagaa ctagtatect gagagtgccg gagaagagcc tgtttggcac  
 61 agccatgacc cggcaagcag ggagctcttg gctcctccgt ggtctcctgc tcttcgcatt  
 121 gtctgcttcc ggcgtcgtc ccttcaactg ggtatctccg gagccccgca gccgagcaag  
 181 caagattcga gtgcaccctc ggggcaacct ctgggagacc ggtcacttca tgggcaagaa  
 241 gagcctggaa cccccgagcc tgctactggg ggggacagca cccctaaca ccccgaggga  
 40 301 ccagagacta cagctgagtc atgatctgct caggatcctc ctgcgaaaga aagctctagg  
 361 catgaacttc agtggcccag ctcccccaat ccagtacagg aggtgctgag agccactact  
 421 gcagaagtga tgccaataat ggaacaaacc ggatgctggg cttagaatgt gtccatccag  
 481 ggaagctgac aatggaaccc tagcagtggc ctctctgga tgtaaatcct aagctcaaat  
 541 gtgttaactc gttactgtga tttctgggtg ggtcaccagg aatgttaatg atgcagacac

601 aaagtctttc ctgctgtatt tcttgcttcc ctggtgaagt ggtgaataaa aatactctct  
661 tc

SEQ ID NO: 184

Amino acid sequence of mouse NMB encoded by the DNA sequence shown in SEQ ID NO:  
5 183.

MTRQAGSSWLLRGLLLFALFASGVAPFNWDLPEPRSRASKIRVHPRGNLWATGHFMGKKS  
LEPPSLSLVGTAPPNTPRDQRLQLSHDLLRILLRKKALGMNFSGPAPPIQYRRLLEPLLQ  
K

SEQ ID NO: 185

10 ENSRNOESTT00000034038 cDNA sequence, EnsEMBL transcript [Rattus norvegicus]

1 cagagaaaac ccgtttggca cagccatgac cgggcaagca gggagcactt ggctgctccg  
61 tgggtctcctg ctctttgcct tggtcgtttc cggcatcacc cccttttccct gggatctccc  
121 ggagccccgc agccgagcaa gcaagattcg agtgcaccct cggggcaacc tctgggcgac  
181 tggtcacttc atgggcaaga agagtctgga acccccgagc ctgtcactgg tgggtacagc  
15 241 accccctatc acccagaggg agcagagact acagctgagt catgatctgc tcaggatcct  
301 cctgctacag aaagcgctag gcatgaacct cagtgggtcca gctccccaa tccagtacag  
361 gaggtctgctg cagaagtgac gccataatg gaacaaacac ggcactgggc gtagcatgag  
421 tccatccagg gaaactgagg aatggaacct tagcagtggc ctctctgga tgtgaaacct  
481 aagctcaaat gtgttactct gttactgtga tttctggtg ggtcaccagg aatgttaacg  
20 541 atacagacac agtctttcct gctgtatttc ttgcttcctt ggtgaagttg tgaataaaaa  
601 tactgctctt tacgaaga

SEQ ID NO: 186

Amino acid sequence of rat NMB encoded by the DNA sequence shown in SEQ ID NO: 185.

25 MTRQAGSTWLLRGLLLFALFVSGITPFSWDLPEPRSRASKIRVHPRGNLWATGHFMGKKS  
LEPPSLSLVGTAPPITQREQRLQLSHDLLRILLQKALGMNLSGPAPPIQYRRLQK

SEQ ID NO: 187

gi|32307134|ref|NM\_005386.2| Homo sapiens neuronatin (NNAT), transcript variant 1,  
mRNA

1 taggtggcgg gcgggtactt aaggcgcggc caccgcggct gcggcagtg ccccaacagc  
30 61 ggactccgag accagcggat ctcggaacac cctctttctc gaccaccac ctaccattct  
121 tggaaaccatg gcggcagtg cggcggcctc ggctgaactg ctcacatcg gctggtacat  
181 cttccgcgtg ctgctgcagg tggtcctgga atgctgcatt tactgggtag gattcgcttt  
241 tcgaaatcct ccaggagacac agccattgac gagaagttag gtgttcaggt actccctgca  
301 gaagctggca tacacggtgt cgcggaccgg gcggcagtg ttgggggagc gcaggcagcg  
35 361 agccccaac tgaggccca gctcccagcc ctgggcggcc gtatcatcag gtgctcctgt  
421 gcatctcggc cagcacggga gccagtgcg cgcaggaaatg tgggttcccc tgtgttccct  
481 cgccagagga gcaactggca aggtcagtag ggggccagta gaccccgga gaagcagtag  
541 cgacaatgac gaagatacca gatcccttc caacccttt gcaccggtcc cactaagggg  
601 cagggtcgag agaggagggg ggataggggg agcaga'tccc tgagatctgg gcataggcac  
40 661 cgattctga tctggacaaa gtcgggacag caccatccca gcccgaagc cagggccatg  
721 ccagcaggcc ccaccatgga aatcaaaaca ccgcaccagc cagcagaatg gacattctga  
781 catcgccagc cgacgcctg aatcttggtg cagcaccaac cgcgtgcctg tgtggcggga  
841 ctggaggggca cagttgagga aggaggggtg ttaagaaata cagtggggcc ctctcgctgt  
901 cccttgccca gggcacttgc attccagcct cgtgcattt gctctctcga ttcccccttc

961 ctcctcactg cctcccaagc ccaccctact ccaaaataat gtgtcacttg atttggaact  
 1021 attcaagcag taaaagtaaa tgaatccac ctttactaaa acactttctc tgaaccccc  
 1081 ttgccctca ctgactctgc tttccctgg tctcatgcag ttgtgggtcaa tattgtggta  
 1141 atcgctaatt gtactgattg ttttaagtgt cattagttgt gtctccccag ctagattgta  
 5 1201 agctcctgga ggacaggac caccctctaca aaaaataaaa aaagtacctc cctgtctcgc  
 1261 cacagtgtcc caggaccctg cgggtgcagta gaggcgcacc aaaaaaaaaa aaaaaaaaaa  
 1321 aaaaaaaaaa aaaaaaaaaa

SEQ ID NO: 188

Amino acid sequence of human NNAT encoded by the DNA sequence shown in SEQ ID NO: 187.

MAAVAAASAELLIIGWYIFRVLLQVFLECCIIYWVGFAFRNPPGTQPIARSEVFRYSLQKL  
 AYTVSRTGRQVLGERRQAPN

SEQ ID NO: 189

gi|32307135|ref|NM\_181689.1| Homo sapiens neuronatin (NNAT), transcript variant 2, mRNA

1 taggtggcgg gcgggtactt aaggcgcggc caccgcggct gcggcagtgc gcccaacagc  
 61 ggactccgag accagcggat ctcggaacac cctctttctc gaccaccac ctaccattct  
 121 tggaaccatg gcggcagtgg cggcgccctc ggctgaactg ctcacatcg gctggtacat  
 181 cttccgcgtg ctgctgcagg tggtcaggta ctccctgcag aagctggcat acacggtgtc  
 20 241 gcggaccggg cggcagggtg tgggggagcg caggcagcga gcccacaact gaggccccag  
 301 ctcccagccc tgggcggccg tatcatcagg tgctcctgtg catctcggcc agcacgggag  
 361 ccagtgccgc gcaggaatgt ggggtccctc gtgttccctc gccagaggag cacttggcaa  
 421 ggtcagtgag gggccagtag acccccggag aagcagtacc gacaatgacg aagataccag  
 481 atcccttccc aacccttttg caccggtccc actaaggggc aggggtcgaga gaggaggggg  
 25 541 gataggggga gcagaccctt gagatctggg cataggcacc gcattctgat ctggacaaag  
 601 tcgggacagc accatcccag ccccgaagcc agggccatgc cagcaggccc caccatggaa  
 661 atcaaaacac cgcaccagcc agcagaatgg acattctgac atcgccagcc gacgccctga  
 721 atcttgggtg agcaccaacc gcgtgcctgt gtggcgggac tggagggcac agttgaggaa  
 781 ggagggtggg taagaaatac agtggggccc tctcgtgtgc ccttgcccag ggcacttgca  
 30 841 ttccagcctc gctgcatttg ctctctcgat tcccctttcc tccctactgc ctcccagcc  
 901 caccctactc caaaataatg tgtcacttga tttggaacta ttcaagcagt aaaagtaaat  
 961 gaatccacc tttactaaa cactttctct gaacccccct tgcccctcac tgatcttgc  
 1021 tttccctggg ctcatgcagt tgtgtgcaat attgtggtaa tcgtaattg tactgattgt  
 1081 ttaagtgtgc attagttgtg tctccccagc tagattgtaa gtcctggag gacagggacc  
 35 1141 acctctacaa aaaaataaaa aagtacctc cctgtctcgc acagtgtccc aggaccctgc  
 1201 ggtgcagtag aggcgcacca aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa

SEQ ID NO: 190

Amino acid sequence of human NNAT variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 189.

MAAVAAASAELLIIGWYIFRVLLQVFRYSLQKLAYTVSRTGRQVLGERRQAPN

SEQ ID NO: 191

gi|6754863|ref|NM\_010923.1| Mus musculus neuronatin (Nnat), transcript

1 variant1,m rnagcggact ccgagaccag tagacctcgg cgaacccttg ctctcgacca

```

    61 cccaccact ttcggaacca tggccgcagt ggcagcagcc tcggcagaac tgctcatcat
    121 cggctggtac atcttccgcg tgcgtctgca ggtgttcctg gaatgctgca tttactgggt
    181 aggattcgct tttcgaaatc ctccagggac acagcccatt gcgagaagtg aggtgttcag
    241 gtactccctg cagaagctgg cgcacacggt gtcccggacc gggcggcagg tgctggggga
5   301 gcgcaggcag cgagccccc actgagggcc cagctcccag ccctgggcgg ccgtgtcatc
    361 aggtgtctct gtgcttctcg accagcatgg gagccagtgc cgcgcaggaa tgggggggtcc
    421 cctgtgttcc ctcgtcagag gagcacttgc caaggtcagt gaggggcccgg taggcccccc
    481 agaaaagcag caccgacaat gatgaagata tcagttccct tcccagcccc tttgcccctg
    541 tcccactacc ggcgggtggg agaggagggg ggaagagggg agcaaccctc gagatatggg
    601 cgtaggcacc acattctgat ctggaccaag tcggaacagc accatctcag ccgcacaaga
    661 tcctaccatg aagatcgaac agcccataa ccagcagaat ggacattctg acatcaccag
    721 ctgaagccct acatctcggt gcagaagaga aagtgtcaac tgtgtgcagc atggggggag
    781 tggagggcgt ggggtgggtgga ggaagagggg taagaaaact agtggggccc ccttgctgcc
    841 ccttgccctat ggcacacata ttctgcctt gctccctcat tcccccttt cccccgcct
    901 tcgaaagccc tccccaaaat gtgtcacttg atttgatat attcaaccag taattgaatc
    961 ccacctttac caaaacacgt tctctaacce ccggcccttc actgatcttg cttatccctg
    1021 gtctcacgca gcagttgtgg tcaatattgt ggtagtcgct aattgtactg gtttaagtgt
    1081 gcattagtag tgtctcccca gctagattgt aagctcctgg agacagggac cacctccacc
    1141 aaaaataaaa aaatggacct ctctgtctt gtagtgtcct aggaccctgc agggca
20

```

SEQ ID NO: 192

Amino acid sequence of mouse NNAT encoded by the DNA sequence shown in SEQ ID NO: 191.

```

25  MAAVAAASAELLIIGWYIFRVLLQVFLECCIIYWVGFAFRNPPGTQPIARSEVFRYSLQKL
    AHTVSRGTGRQVLGERRQAPN

```

SEQ ID NO: 193

gi|16758385|ref|NM\_053601.1| Rattus norvegicus neuronatin (Nnat), transcript variant 1, mRNA

```

    1 gcgaaccctt gctctcgacc acccaccac tttcggaacc atggccgcag tggcagcagc
    61 ctccgcagaa ctgctcatca tcggctggta catcttccgc gtgctgctgc aggtgttcct
    121 ggaatgctgc atttactggg taggattcgc ttttcgaaat cctccaggga cacagcccat
    181 tgcgagaagt gaggtgttca ggtactccct gcagaagctg gcgcacacgg tgtcccggac
    241 cgggcggcag gtgctggggg agcgcaggca ccgagcccc aactgaggcc ccattctcca
    301 gccctgggag gccgtgtcat caggtgtctc tgtgtctctc gaccagcatg ggagccaatg
    35  361 ccgcgcagga atgggggggtc ccctgtgtct cctcgtcaga ggagcacttg ccaaggctcag
    421 tgagggggcg gtaggtcccc agaaaagcag caccgacaat gatgaagaca tcagttcctt
    481 tcccagcccc ccccccttt gccctgtcc catggccggc ggggtgggaga ggatggggga
    541 agagggggagc aaccctcgag atatgggcgt aggcaccaca ttctgatctg gaccaagttg
    601 gaacagcacc atctcagccg cacagatcct accatggaga gctaacaccc caccaaccag
    661 cagaatggac attctgacat caccagctga aaccctgaat ctcggtgcag aagagaaaagt
    721 gtcaactgcg tgcagcactg ggggagtgga ggggtgtggg ggtggaggaa gaggggttaag
    781 aaaactagtg ggccctcttt gctgtccctt gcctatggca cgcataattc tgccctgtct
    841 cctcactccc cctctcccct gccttccaaa gccccacccc ccaaaaatgt gtcacttgat
    901 ctggacctat tcaaccagta attggatccc acctttacca aaacaccgtc tctgaccccc
    45  961 ggcccttcac tgatcttgct tatccctggg ctcacgcagc agttgtgggt gctattgtgg
    1021 tagtcgctaa ttgtactagt ttacgtgtgc attagtgtg tctccccggc tagattgtaa
    1081 gtccttgag acagggacca cctccacaaa aaataaaaaa acggacctct cctgtcttgt
    1141 agtgtgctag gaccctgcag ggcagtgggg gtgcacca

```

SEQ ID NO: 194

Amino acid sequence of rat NNAT encoded by the DNA sequence shown in SEQ ID NO: 193.

MAVAAASAELLIIGWYIFRVLLQVFLECCYVWGFARNPPTQPIARSEVFRYSLQKL  
AHTVSRTRGRQVLGERRHAPN

5 SEQ ID NO: 195

gi|17157992|ref|NM\_058164.1| Homo sapiens olfactomedin 2 (OLFM2), mRNA

```
1 ctaggtcggg acgccgggct aggggcgcgt catgtggcgg ctcacgggtcc cgccgccgct
61 gctgctgctg ctgtgctcag gcctggccgg acagactctc ttccagaacc cagaagagggg
121 ctggcagctg tacacctcag cccaggcccc tgacgggaaa tgcattctga cggccgtgat
10 181 cccagcgcag agtacctgct ctcgagatgg caggagtcgg gagctgcggc aactgatgga
241 gaaggtccag aacgtctccc agtccatgga ggtccttgag ttgcccagct atcgcgacct
301 ccagtatgta cgcggcatgg agaccctcat gcggagcctg gatgcgcggc tccgggcagc
361 tgatgggtcc ctctcggcca agagcttcca ggagctgaag gacaggatga cggaaactgtt
421 gcccctgagc tcggctcctgg agcagtacaa ggcagacacg cggaccattg tacgcttgcg
15 481 ggaggagggtg aggaatctct ccggcagctc ggccggccatt caggaggaga tgggtgccta
541 cgggtatgag gacctgcagc aacgggtgat ggccctggag gcccggtccc acgcctgcgc
601 ccagaagctg ggctgtggga agctgaccgg ggtcagtaac cccatcaccg ttcgggccat
661 ggggtcccgc ttcggctcct ggatgactga cagatggcc cccagtgcgg atagccgggt
721 ctggtacatg gatggctatt acaaagccg ccgggtcctg gaggttccgta ccctgggaga
20 781 cttcatcaaa ggccagaact ttatccagca cctgctgccc cagccgtggg cgggcacggg
841 ccacgtggtg tacaacggct ccctgttcta taacaagtac cagagcaacg tgggtggtcaa
901 ataccacttc cgctcgcgct ctgtgctggt gcagaggagc ctcccgggcg ccggttacaa
961 caacaccttc ccctactcct ggggcggctt ctccgacatg gacttcatgg tggacgagag
1021 cgggctctgg gctgtgtaca ccaccaacca gaacgcgggc aacatcgtgg tcagccggct
25 1081 ggaccgcgac accctcgagg tcatgcggtc ctgggacacc ggctacccca agcgcagcgc
1141 tggcgaggcc ttcattgatc gcgggtgtgt ctacgtgacc aactcccacc tggctggggc
1201 caagctctac ttgcctatt ttaccaacac gtccagttac gactacacgg acgtgccctt
1261 ccacaaccag tattcccaca tctcgatgct ggattacaac ccccgggagc gcgcccctta
1321 tacctggaac aacggccacc aggtgtctta caatgtcacc ctgtttcacg tcatcagcac
30 1381 ctctggggac ccctgagcca atgctgtggc tcgggctgct gcctgggggg cctccggggg
1441 ctggggggccc ttttcttctt gcctgtgtcc ctcaagggtg atctctctgt ctctgtcacg
1501 ccttttcttc ccgccttttt gctgggcttt tgttctctgc ctatgtattt ctgtctattt
1561 tttcaatttc cctctctctc ttttattgat ctctgctttt aatacaccac tttcttcttt
1621 ctgccttttt atggatgtct ttttcttttt atggctctgg ttctccagtt tttccgctt
35 1681 ctgcctctct ctgtctctct ctctctgtcc ttccaccctt cctccttgc tttccacca
1741 ttcctcatcc ctactccca cccccacccc cccccccagg agttgagtgc atggatctgt
1801 tttctttttt atttacactt tttctttccg gtttgccgga ataaacagga cctttgacat
1861 ttgaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa
```

SEQ ID NO: 196

40 Amino acid sequence of human OLFM2 encoded by the DNA sequence shown in SEQ ID NO: 195.

```
11 MWPLTVPPPLLLLLCSGLAGQTLFQNPEEGWQLYTSAPDQKICITAVIPAQSTCSRDG
12 RSRELRLQMEKVQNVQSMEVLELRITYRDLQYVRGMETLMRSLDARLRAADGSLAKSFQ
13 ELKDRMTLELLPLSSVLEQYKADTRTIVRLREEVRNLGSLAAIQEEMGAYGYEDLQQRVM
45 14 AEARLHACAQKLGCGKLTGVSNPITVRAMGSRFGSWMTDTMAPSADSRVWYMDGYKGR
15 RVLEFRTLGDPIKQNFIOHLLPQPWAGTGHVVYNGSLFYNKYQSNVVVKYHFRSRSVLV
16 QRSLPGAGYNNTPFYSWGGFSDMDFMVDESGLWAVYTTNQNAGNIVVSRDPHTLEVMSR
17 WDTGYPKRSAEAFMICGVLYVTNSHLGAKVYFAYFTNTSSYEYTDVPPHNQYSHISML
18 DYNPRERALYTNNGHQVLYNVTLFHVISTSGDP
```

## SEQ ID NO: 197

Amino acid sequence of human OLFM2, a soluble active secreted form derived from SEQ ID NO:196.

```

5  QTLFQNPEEGWQLYTSAQAPDGKCICTAVIPAQSTCSRDRSRELRLQMEKVQNVQSME
   VLELRITYRDLQYVRGMETLMRSLDARLRAADGSLSAKSFQELKDRMTELLPLSSVLEQYK
   ADTRTIVRLREEVRNLGGSLAAIQEEMGAYGYEDLQQRVMALEARLHACAQKLGCGKLTG
   VSNPITVRAMGSRFGSWMTDTMAPSADSRVWYMDGYKGRRVLEFRTLGDFIKGQNFQIH
   LLPQPWAGTGHVVYNGSLFYNKYQSNVVVKYHFRSRSVLVQSRSLPGAGYNNTFPYSWGGF
10  SDMDFMVDESGLWAVYTTNQNAGNIVVSRLDPHTLEVMSWDGTGYPKRSAGEAFMICGVL
   YVTNSHLAGAKVYFAYFTNTSSYEYTDVPPHNQYSHISMLDYNPRERALYTNWNGHQVLY
   NVTLFHVISTSGDP

```

## SEQ ID NO: 198

gi|31343362|ref|NM\_173777.2| Mus musculus hypothetical protein A030009A06 (A030009A06), mRNA

```

15  1  gagttgagcg gcaggctacc tctggcttct tacacatacc tgacccatca ggcccccaga
   61  accactttac agatgaggaa actgagacag accggaacaa ctattgctgg aggtcagact
   121  ctcttccaga gcccggagga gggctggcag ctttatacgt cagcccaggc acctgatggc
   181  aagtgcgtct gcacagccgt gatccctgcg cagagcacct gtgcccgaga cggtcggagc
   241  agagagcttc ggcaactcat ggagaaggtc cagaatgtgt cccagtccat ggaggtcctt
20  301  gagctaagga cattccggga tctccagtat gttcgcagca tggagaccct catgctggagc
   361  ctggatgcaa ggctcagggc agccgatggg tcagtctcag ccaaaagctt tcaggaactg
   421  aaggacagga tgacagagct gctgcccttg agttcagtgc tggagcagta caaagcagac
   481  acacgaacca ttgtgcgctt gcgggaggag gtgaggaacc tctctggcaa cttggctgcc
   541  atccaagagg aaatgggtgc ctacgggtac gaggacttgc aacagcgctg gatggccctg
25  601  gaagcccgac tccatgcctg cgcgcagaag ctgggctgcg ggaagctgac aggcgtcagt
   661  aaccctatta ccatccgggc catggggtcc cgcttcggtt cctggatgac tgacacaatg
   721  gccccagtg cagacagcag ggtctgttac atggatggtt attacaaagg ccgccgagtg
   781  ctggagttcc gtactctggg agacttcata aaaggccaga acttcacca gcacctgctg
   841  ccacagccat gggcaggtac gggccatgtg gtatacaatg gctctctctt ctacaataag
30  901  taccagagca atgtggtggt caagtaccac ttccggtccc gtcggtgct ggtgcagagg
   961  agcctcccgg gggctgggta caataaacac ttcccctatt cctggggcgg cttctcggac
   1021  atggacttca tggtagacga gagtgggctg tgggctgtgt ataccacca ccagaatgctg
   1081  ggcaacatcg tagtcagtcg gctggaccct cacaccctgg aggtcgtgag atcctgggac
   1141  accgggtacc ctaagcgcag cgccggcgag gccttcata tctgcggtgt cctctatgtg
35  1201  accaactctc acctggccgg agccaaggtc tactttgcgt acttcacca cactccagc
   1261  tatgagtaca cggatgtgcc cttcaacaac cagtattgcg acatctcgat gctggattac
   1321  aaccccgagg agcggggcct gtatacctgg aacaacgggc accagtgct gtacaacgtc
   1381  accctgttcc acgtcatcag cactgcccgg gacccttagg tgcccctgca agggcttgg
   1441  ggagccttcc cacatgcgct gtgaccccca cccagcctt ctcttggtca tgcccttgcc
40  1501  ttcttagatt cegtctcccc acttccccag cccagctttc tgttctcagt atctctacce
   1561  atgcatttcc cccattttat tgatctctgc ttttgatact ccacttctga gtctttctgc
   1621  ctttttatgg atgcttctct tccttattga tcgaacctct cttctctct cctccatct
   1681  actctccctc ctcttccttt cccactttcc cacattcctt acccctcact cccaccctat
   1741  ccctggatct gagtgcattg attttgttt ttaaatttat tatttacatg ttttctttcc
45  1801  ggggtgccag aataaacggg accttt

```

## SEQ ID NO: 199

Amino acid sequence of mouse OLFM2 encoded by the DNA sequence shown in SEQ ID NO: 198.

MRKLRQTGTTIAGGQTLFQSPPEEGWQLYTSAQAPDGKCVCTAVIPAQSTCARDGRSRELRL

QLMEKVQNVQSMEVLELRTFRDLQYVRSMETLMRSLDARLRAADGSVSAKSFQELKDRM  
 TELLPLSSVLEQYKADTRTIVRLREEVRNLSGNLAAIQEEMGAYGYEDLQQRVMALEARL  
 HACAQKLGCGKLTGVSNPITIRAMGSRFGSWMTDTMAPSADSRVWYMDGYKGRVLEFR  
 TLGDFIKGQNFIOHLLPQPWAGTGHVVYNGSLFYNKYQSNVVVKYHFRSRSVLVQSRSLPG  
 5 AGYNNTFPYSWGGFSDMDFMVDESGLWAVYTTNQNAGNIVVSRLDPHTLEVVRSWDTGYP  
 KRSAGEAFMICGVLYVTNSHLAGAKVYFAYFTNTSSYEYTDVPFNNQYSHISMLDYNPRE  
 RALYTWNNGHQVLYNVTLFHVISTAGDP

SEQ ID NO: 200

10 gi|34860388|ref|XM\_233742.2| Rattus norvegicus similar to olfactomedin 2; neuronal  
 olfactomedin related ER localized protein 2; noelin 2 (LOC313783), mRNA

1 ccccatcccc atccccaccg ccaccgcccc cttggaccgg gtcattctcg catcgggacc  
 61 ccgacccctc caccacctg ggggtggatg cgacatcttg agtgccgagg aaaggggacg  
 121 gaacacctgc tcttttaggg tgggtggtgcg gggcatgagg cgaggggagc gatgtcgggtg  
 181 ccgctgctca agatcggggc ggtgctgagc accatggcca tggctactaa ctggatgtcg  
 15 241 cagacgctgc cctcgctcgt gggactcaac agcaccgtgt cccgcgcggg ctccctcagag  
 301 aaaatcactc tcttccagag ccagaagag ggctggcagc tgtacacgtc agcgcaggcg  
 361 ccgcatggca aatgcactcg cacagctgtg attcctgcac agagcacttg tgcccagatg  
 421 ggtcggagca gagagcttcg ccagctcatg gagaagggtc agaattgtgc tcagtccatg  
 481 gaagtccttg agctaaggac ataccgggat ctccagtatg ttcgcagcat ggagaccctc  
 20 541 atgcggagcc tggatgcaag gtcaggaca gccgatgggt cagtctcagc caaaagcttt  
 601 caggaactga aggacaggat gacagagctg ctgcccctga gttcagtgtc ggagcagtac  
 661 aaagcagaca caccgacat tgtgcgcctg cgggaggagg tgaggaaact ctcaggcaac  
 721 ctggctgcca tccaggagga aatgggtgct tacgggtacg aggacttgca gcagcgcgtg  
 781 atggccctgg agggccgact ccatgcctgc gcgcagaagc tgggctgcgg gaagctgaca  
 25 841 ggcgtcagta accccattac catccgagcc atgggatccc gcttcggctc ctggatgact  
 901 gacacgatgg cccccagtgc agacagtcgg gtgagtgacc ccaactccagt gactgcctgc  
 961 aggtgttcag tgaacattct tagcccagtg tccagtgaat ctgtgatccc agtctggtac  
 1021 atggatgggt attacaaagg ccgccgagtg ctggagttcc gtactttggg agacttcac  
 1081 aaaggccaga acttcaccca gcacctgctg ccgcagccat gggcaggtag gggccatgtg  
 30 1141 gtatacaatg gatctctgtt ttacaacaag taccagagca atgtggtggt taagtaccac  
 1201 ttccggtccc gctccgtggt ggtgcagagg agcctcccgg gggctgggta caataacacc  
 1261 tccccctatt cctggggcgg cttctcggac atggacttca tggtagacga gagcgggctg  
 1321 tgggctgtgt ataccaccaa ccagaacgag ggcaacattg tggtcagccg gctggaccct  
 1381 cacaccctgg aggtcgtgag gtctcgggac actgggtacc ctaagcgcag cgcgggtgag  
 35 1441 gccttcattg tctgtggtgt cctctacgtg accaactctc acctggccgg agccaaggtc  
 1501 tactttgctg acttcaccaa cacgtccagc tatgagtaca cggatgtgcc cttccacaac  
 1561 cagtactcgc acatctccat gctggattac aaccccaggg agcgggcccct gtatacctgg  
 1621 aacaacgggc accagggtgct gtacaacgct accctcttcc acgtcatcag cactgccggg  
 1681 gacccttag

40 SEQ ID NO: 201

Amino acid sequence of rat OLFM2 encoded by the DNA sequence shown in SEQ ID NO:  
 200.

MSVPLLKIGAVLSTAMVTVNWSQTLPSLVGLNSTVSRAGSSEKITLPQSPEEGWQLYTS  
 45 AQAPDGKICITAVIPAQSTCARDGRSRELRLMEKVQNVQSMEVLELRTYRDLQYVRSM  
 ETLMRSLDARLRTADGSVSAKSFQELKDRMTELLPLSSVLEQYKADTRTIVRLREEVRNL  
 SGNLAAIQEEMGAYGYEDLQQRVMALEARLHACAQKLGCGKLTGVSNPITIRAMGSRFGS  
 WMTDTMAPSADSRVSDPTPVTACRCSVNILSPVSSESVIPVWYMDGYKGRVLEFRTL  
 DFIKQNFIOHLLPQPWAGTGHVVYNGSLFYNKYQSNVVVKYHFRSRSVLVQSRSLPGAGY  
 NNTFPYSWGGFSDMDFMVDESGLWAVYTTNQNAGNIVVSRLDPHTLEVVRSWDTGYPKRS  
 50 AGEAFMICGVLYVTNSHLAGAKVYFAYFTNTSSYEYTDVPFNNQYSHISMLDYNPRER  
 YTWNNGHQVLYNVTLFHVISTAGDP



## SEQ ID NO: 202

gi|7657070|ref|NM\_014322.1| Homo sapiens opsin 3 (encephalopsin, panopsin) (OPN3), mRNA

```
1  cgagccccgc cgcaagctga ggcctccgc cgcagggcg cgccggcgcc gggccatgta
5  61  ctccggggaac cgcagcgggc gccacggcta ctgggacggc ggcggggccg cgggcgctga
121 ggggcccggcg ccggcgggga cactgagccc cgcgcccctc ttcagccccg gcacctacga
181 gcgcctggcg ctgctgctgg gctccattgg gctgctgggc gtcggcaaca acctgctggg
241 gctcgtcttc tactacaagt tccagcggct ccgcactccc actcacctcc tcttggtcaa
301 catcacatatt agcgacctgc ttggtgccct cttcggggtc acctttacct tcgtgtcctg
10 361 cctgaggaac ggctgggtgt gggacacagt gggctgctgt tgggacgggt ttagcggcag
421 cctcttcggg attgtttcca ttgccacct aaccgtgctg gcctatgaac gttacattcg
481 cgtggtccat gccagagtga tcaatttttc ctgggcctgg agggccatta cctacatctg
541 gctctactca ctggcggtgg caggagcacc tctcctggga tggacagggt acatcctgga
601 cgtacacgga ctaggctgca ctgtggactg gaaatccaag gatgccaacg attcctcctt
15 661 tgtgcttttc ttatttcttg gctgcctggg ggtgcccctg ggtgtcatag cccattgcta
721 tggccatatt ctatatcca ttcgatcttc tcgtgtgtgt gaagatcttc agacaattca
781 agtgatcaag attttaaaat atgaaaagaa actggccaaa atgtgctttt taatgatatt
841 caccttctct gtctgttggg tgccttatat cgtgatctgc ttcttggtgg ttaatggtca
901 tggtcacctg gtcactccaa caatatctat tgtttcgtac ctctttgcta aatcgaacac
20 961 tgtatacaat ccagtgtatt atgtcttcat gatcagaaag ttctgaagat cccttttgca
1021 gcttctgtgc ctccgactgc tgagggtgcc gaggcctgct aaagacctac cagcagctgg
1081 aagtgaatg cagatcagac ccattgtgat gtcacagaaa gatggggaca ggccaaagaa
1141 aaaagtgact ttcaactctt cttccatcat ttttatcatc accagtgatg aatcactgtc
1201 agttgacgac agcgacaaaa ccattggggg ccaaagtgtg atgttaatcc aagttcgtcc
25 1261 tttgtaggaa tgaaggatgg caacgaaagg tggggcctta aattggatgc cacttttggg
1321 ctttcatcat cctcctgaag aagaagtgtc tgggaataccc gttctatgta atatcaacag
1381 aaccttgtgg tccagcagga aatccgaatt gcccatatgc tcttgggcct caggaagagg
1441 ttgaacaaaa acaaatctt ttaattcaac ggggtgcttta cataatgaaa aaaccacttg
1501 tgcacacgat gggcatctaa catcatcatc ttctaattgt ttggagattt tcatttcaaa
30 1561 tatattttt aaattactct attttccaaa acacgtaatg catttttctc gaaaatacct
1621 tactgtaaaa ataactgtcg cgtacacatg tgtgaagtag ctagaacata ctgaattttt
1681 tttgtactgt tggactctat tcagtgtcat gtcctatata tgatcaagtt atcaaggaga
1741 taattctaga atgaaaaaga aaatcctctt gttggaaaca aaagacgttt tatatgtgca
1801 gtatgacaaa gaggagtttc agagacaact ttgaatcctt gtcagcctgg agaccagcac
35 1861 cagaggaatc tacaaggcaa actcccatat atttgcttcc cccaaattgc tggccctaca
1921 gactcaaaagc tcttttctct tgttttgttg tttctctaaa aatttactgt tctttgtcga
1981 tgctatataa gccagggagt tctaagacgc cagctctttg agatttgctc attcccctgt
2041 atttcccaaa tatatattac atataccgc taataaattt atgtttgttt taacaaaaaa
2101 aaaaaaaaaa
```

## 40 SEQ ID NO: 203

Amino acid sequence of human OPN3 encoded by the DNA sequence shown in SEQ ID NO: 202.

```
45  MYSGNRSGGHGYWDGGGAAGAE GPAPAGT LSPAPLFSPGTYERLALLLSIGLLGVGNL
LVLVLYYKQRLRTPHLLLVNISLSDLLVSLFGVTFVFSCLRNQWVWDTVGCVWDGFS
GSLFGIVSIATLTVLAYERYIRVVHARVINFSWAWRAITYIWLYSLAWAGAPLLGWNRYI
LDVHGLGCTVDWKS KDANDSSFVFLFLGCLVPLGVIAHCYGHILYSIRMLRCVEDLQT
IQVIKILKYEKKLAKMCF LMIFFLVCWMPYIVICFLVNVNGHGLVPTTISIVSYLFAKS
NTVYNPVIYVFMIRKFRSLLQLLCLRLRLRCQRP AKDLPAAGSEMQIRPIVMSQKQDGRP
KKKVTFNSSSIIFIITSDESLSVDDSDKTIGVQSLMLIQVRPL
```

## 50 SEQ ID NO: 204

(gi|27463268:201-667, 24621-24703, 42570-42821, 45876-47409) Homo sapiens  
encephalopsin splice variant 1 (OPN3) gene, alternatively spliced; and choroideremia-like  
protein (CHML) gene, complete cds

```

5      1 gcggcgaggc cccaacccca cccagtgcgg agcgcgccgc gagccccgcc gcaagctgag
      61 cgccctccgc cgccaggcgc gccggcgccg ggccatgtac tcggggaacc gcagcgccgc
      121 ccacgggtac tgggacggcg gcggggccgc gggcgctgag gggccggcgc cggcggggac
      181 actgagcccc gcgccccctc tcagcccccg cacctacgag cgcctggcgc tgctgctggg
      241 ctccattggg ctgctgggcg tcggcaacaa cctgctgggtg ctgctcctct actacaagtt
      301 ccagatggctc cgcactccca ctccctctct cctgggtcaac atcagcctca gcgacctgct
10     361 ggtgtccctc ttcgggggtca cctttacctt cgtgtcctgc ctgaggaaac gctgggtgtg
      421 ggacaccgtg ggctgcgtgt gggacggggt tagcggcagc ctcttcggtt ggatcagcca
      481 gcttcaggca gccactaggg aagccagagc ctccatgggt ccagtgcagc aaggcactat
      541 ctgcatgcag ctctgtgtgt tggaagatct tcagacaatt caagtgatca agattttaaa
      601 atatgaaaag aaactggcca aaatgtgctt tttaatgata ttcacctctc tggctctgtt
15     661 gatgccttat atcgtgatct gcttcttggt ggtaaatggt catggtcacc tggctactcc
      721 aacaatatct attgttctgt acctcttctc taaatcgaa actgtatata atccagtgt
      781 ttatgtcttc atgatcagaa agtttcgaag atcccttttg cagcttctgt gcctcgcact
      841 gctgaggtgc cagaggcctg ctaaagacct accagcagct ggaagtgaac tgcagatcag
      901 acccattgtg atgtcacaga aagatgggga caggccaaag aaaaaagtga ctttcaactc
20     961 ttcttcocat atttttatca tcaccagtga tgaatcaact tcagttgacg acagcgacaa
      1021 aaccaatggg tccaaagtgt atgtaatcca agttcgtcct ttgtaggaat gaagaatggc
      1081 aacgaagaat ggggccttaa attggatgcc acttttggtg tttcatcata agaagtgtct
      1141 ggaatacccg ttctatgtaa tatcaacaga accttggtgt ccagcaggaa atccgaattg
      1201 cccatatgct cttgggcctc aggaagaggt tgaacaaaaa caaattcttt taattcaacg
25     1261 ggtgctttac ataataaaaa aaccacttgt ggcacacgat gggcatctaa catcatcatc
      1321 ttctaattgt ttggagattt tcatttcaaa tatatttttt aaattactct attttccaaa
      1381 acacgtaatg catttttctc gaaaatacct tactgtaaaa ataactgtcg cgtacacatg
      1441 tgtgaagtag ctagaacata ctgaattttt tttgtactgt tggactctat tcagtgtcat
      1501 gtcctatatc tgatcaagtt atcaaggaga taattctaga atgaaaaaga aaatcctctt
30     1561 gttggaataa aaagacgttt tatattgaca gtatgacaaa gaggagtttc agagacaact
      1621 ttgaatcctt gtacgcctgg agaccagcac cagaggaatc tacaaggcaa actcccatat
      1681 atttgcttcc cccaaattgc tgcccttaca gactcaaagc tctttttctt tgttttgtt
      1741 tttctctaaa aatttactgt tctttgtcga tgctatataa gccagggagt tctaagacgc
      1801 cagctctttg agatttgcct attcccctgt atttcccaca tatatattac atataccgcg
35     1861 taataaattt atgtttgttt ttctctgttc aatctgtctt ttgttatagg ggccccagcc
      1921 aaggaacctc aagtgggtag aaggaaaaat tattttttct ttccctacaa actgaacatg
      1981 gattattaga actcaagggt ttcattgaca atatagaaa gaaacactga atcattttat
      2041 tttattgccc aatttttatt tcttatatga ctctagtgtt tcatcttcat aattaatcat
      2101 gtttgaagga tttctgagtg actcagcagc ctgttaaaga aggatgaacc aaagaaaaca
40     2161 tttcactaaa tgtgctttta aaaaatcaagt gtattgctgg ttctgctgca gtatgtagtc
      2221 gaagaataaa ttagtaaat gcttctgagg gtctgaaatt gaataaagta atggctttgt
      2281 atttctataa aagttgtctc cccttgtttc ctttccattc tggcacatgt agacat

```

SEQ ID NO: 205

45 Amino acid sequence of human OPN3 variant ORF number 1 encoded by the DNA sequence  
shown in SEQ ID NO: 204.

```

      MYSGNRSGGHGYWDGGGAAGAEGPAPAGTLSAPLFSPGTYERLALLLSIGLLGVGNL
      LVLVLYYKFORLTPHLLLVNISLSDLLVSLFGVTFVVSCLRNGWVWDTVGCVWDGFS
      GSLFGWISQLQAATREARASMGFPVQQGTICMQLRCVEDLQTIQVIKILKYEKKLAKMCFL
50     MIFTFLVCWMPYIVICFLVVNGHGLVTPTISIVSYLFAKSNTVYNPVIYVFMIRKFRRS
      LLQLCLRLRLRCORPAKDLPAAGSEMQIRPIVMSQKDGDRPKKKVTFNSSSIIFIITSDE
      SLSVDDSDKTNGSKVDVIQVRPL

```

SEQ ID NO: 206

(gi|27463268:201-667, 35987-36306, 42570-42821, 45876-47409) Homo sapiens encephalopsin splice variant 1 (OPN3) gene, alternatively spliced; and choroideremia-like protein (CHML) gene, complete cds

```

1  gcggcgggagc  cccaacccca  cccagtgcgg  agcgcgcgcg  gagccccgcc  gcaagctgag
5  61  cgcctccgcc  cgccaggcgc  gccggcgccg  ggccatgtac  tcggggaacc  gcagcggcgg
121 ccacggctac  tgggacggcg  gcggggccgc  gggcgctgag  gggccggcgc  cggcggggac
181 actgagcccc  gcgcccctct  tcagcccccg  cacctacgag  cgcctggcgc  tgctgctggg
241 ctccattggg  ctgctgggcg  tcggcaacaa  cctgctggtg  ctgctcctct  actacaagtt
301 ccagcgggctc  cgcactccca  ctcacctcct  cctgggtcaac  atcagcctca  gcgacctgct
10 361 ggtgtccctc  ttccgggtca  cctttacctt  cgtgtcctgc  ctgaggaacg  gctgggtgtg
421 ggacaccgtg  ggctgcgtgt  gggacggggt  tagcggcagc  ctcttcggga  ttgtttccat
481 tgccacccta  accgtgctgg  cctatgaacg  ttacattcgc  gtggtccatg  ccagagtgat
541 caatTTTTCC  tgggcctgga  gggccattac  ctacatctgg  ctctactcac  tggcgtgggc
601 aggagcacct  ctccctggat  ggaacaggta  catcctggac  gtacacggac  taggctgcac
15 661 tgtggactgg  aaatccaagg  atgccaacga  ttccctcctt  gtgcttttct  tatttcttgg
721 ctgcctggtg  gtgcccctgg  gtgtcatagc  ccattgctat  ggccatattc  tatattccat
781 tcgaatgctt  cgttgtgtgg  aagatcttca  gacaattcaa  gtgatcaaga  ttttaaaata
841 tgaaaagaaa  ctggccaaaa  tgtgcttttt  aatgatattc  accttccctg  tctgttggat
901 gccttatatc  gtgatctgct  tcttggtggt  taatggtcac  ggtcacctgg  tcactccaac
20 961 aatatctatt  gtttcgtacc  tctttgctaa  atcgaacact  gtatacaatc  cagtgattta
1021 tgtcttcatg  atcagaaagt  ttccaagatc  ccttttgtag  cttctgtgcc  tccgactgct
1081 gagggtgccg  aggcctgcta  aagacctacc  agcagctgga  agtgaaatgc  agatcagacc
1141 cattgtgatg  tcacagaaag  atggggacag  gccaaagaaa  aaagtgactt  tcaactcttc
1201 ttccatcatt  tttatcatca  ccagtgtatg  atcactgtca  gttgacgaca  gcgacaaaac
25 1261 caatgggtcc  aaagttgatg  taatccaagt  tcgtcctttg  taggaatgaa  gaatggcaac
1321 gaaagatggg  gccttaaatt  ggatgccact  tttggacttt  catcataaga  agtgtctgga
1381 atacccttcc  tatgtaatat  caacagaacc  ttgtggtcca  gcaggaaatc  cgaattgccc
1441 atatgctctt  gggcctcagg  aagagggtga  acaaaaacaa  attcttttaa  ttcaacgggt
1501 gctttacata  atgaaaaaac  cacttgtggc  acacgatggg  catctaactc  catcatcttc
30 1561 taatgtgttg  gagattttca  tttcaaatat  attttttaaa  ttactctatt  ttccaaaaca
1621 cgtaatgcat  ttttctcgaa  aataccttac  tgtaaaaata  actgtcgcgt  acacatgtgt
1681 gaagtagcta  gaacatactg  aatttttttt  gtactgttgg  actctattca  gtgtcatgtc
1741 ctatatctga  tcaagttatc  aaggagataa  ttctagaatg  aaaaagaaaa  tcctcttggt
1801 ggaaacaaaa  gacgttttat  atgtgcagta  tgacaaagag  gagtttcaga  gacaactttg
35 1861 aatccttgtc  agcctggaga  ccagcaccag  aggaatctac  aaggcaaact  cccatatatt
1921 tgcttcccc  aaattgctgc  ccctacagag  tcaaagctct  ttttctttgt  tttgttggtt
1981 ctctaaaaat  ttactgttct  ttgtcgatgc  tatataagcc  agggagtctc  aagacgccag
2041 ctctttgaga  tttgctcatt  cccctgtatt  tcccacatat  atattacata  taccgcgtaa
2101 taaatttatg  tttgtttttc  tcttgtcaat  ctgtcttttg  ttataggggc  cccagccaag
40 2161 gaacctaaag  tgggtagaag  gaaaaattat  tttttcttcc  cctacaaact  gaacatggat
2221 tattagaact  caaggttttc  attgacaata  tagaaaagaa  acactgaatc  attttatttt
2281 attgcecaat  ttttatttct  tatatgactc  tagtgtttca  tcttcataat  taatcatgtt
2341 tgaaggattt  ctgagtgact  cagcagcctg  ttaaagaagg  atgaaccaa  gaaaacattt
2401 cactaaatgt  gcttttaaaa  atcaagtgtg  ttgctgggtc  tgctgcagta  tgtagtcgaa
45 2461 gaataaatta  gtaaatgtct  tctgagggtc  tgaaattgaa  taaagtaatg  gctttgtatt
2521 tctataaaa  ttgtctcccc  ttgtttcctt  tccattctgg  cacatgtaga  cat

```

SEQ ID NO: 207

50 Amino acid sequence of human OPN3 variant ORF number 2 encoded by the DNA sequence shown in SEQ ID NO: 206.

```

55  MYSGNRSGGHGYWDGGGAAGAEGPAPAGTSLSPAPLFSPGTYERLALLLGSIGLLGVGNL
    LVLVLYYKFORLRTPTHLNLSLSDLLVSLFGVTFVFSCLRNGWVWDTVGCVWDGFS
    GSLFGIVSIATLTVLAYERYIRVVHARVINFSWAWRAITYIWLYSLAWAGAPLLGWNRYI
    LDVHGLGCTVDWKSNDANDSSFLFLFLGCLVPLGVIAHCYGHILYSIRMLRCVEDLQT
    IQVIKILKYEKKLAKMCFMIFTFLVCWMPYIVICFLVNGHGLVPTISIVSYLFAKS

```

NTVYNPVIYVFMIRKFRSLLQLLCLRLRLRCQRPALDPAAGSEMQRPIVMSQKDGDRP  
KKKVTFNSSSIIFIITSDESLSVDDSDKTNGSKVDVIQVRPL

SEQ ID NO: 208

gi|6753709|ref|NM\_010098.1| Mus musculus opsin (encephalopsin) (Opn3), mRNA

```

5      1 tcggggccgcc cggtccaccg agccctctct ctcacggcgc gccggcgctcg ggccatgtac
      61 tcgggggaacc gtagtggcga ccagggctac tgggaggacg gggcgggcgc cgagggcgca
      121 gcaccggcgg gcacgcggag ccccgccct ctcttcagcc ccaccgcgta cgagcgccctg
      181 gcgctgctac tcggctgcct cgcgctgctg ggcgtcggcg gcaacctgct ggtgctgctt
      241 ctctactcca agttcccag actgcgcacg cccaccacc tcttcctggt caacctgagc
10     301 ctgggcgacc tgctggtatc cctgttcgga gtcacctca ccttcgcgtc gtgcctgcgg
      361 aacggctggg tgtgggacgc cgtgggctgc gcgtgggacg ggtttagcgg cagcctcttt
      421 gggttcgttt ccattaccac cctcactgtg ctggcctatg aacgttatat ccgtgtggtta
      481 catgccagag tgatcaactt ttcctgggcc tggagggcca ttacctatat ctggctctac
      541 tccttggcat gggcaggagc acctctcctg ggctggaaca ggtacatcct agacatacat
15     601 ggactgggct gtaccgtgga ctggagatcc aaggatgcca acgactcttc ctttgtgctc
      661 ttcctgtttc tcggctgcct ggtggtgccc gtgggcatca tagccattg caacggccac
      721 attctctatt ctgttcgaat gcttcgctgt gttgaagatc ttcagaccat tcaagtgatc
      781 aagatgctaa gatatgaaaa gaaagtagca aagatgtgct ttttgatggc ctttgtcttt
      841 ctcacctgct ggatgcctta cattgtgacc cgcttcttgg tggatcaatg ctatggacac
20     901 ctggtcaccc caactgtgtc tattgtttct tatctctttg ctaaactcgag cactgtgtac
      961 aaccagtta tctacatct catgaacaga aagtttcgga ggtcccttct gcagctccta
     1021 tgcttcggcc tgctgagatg ccagcggcct gctaaaaacc tcccagcggc tgagagtga
     1081 atgcacatca ggcccatcgt gatgtcacag aaagatgggg acaggccaaa gaagaaagt
     1141 acctttaact ctctctctat catctttatc atcaccagtg atgagtcct gtcagtcgag
25     1201 gacagtgaac gaagcagcgc atctaaggtc gatgtcatcc aagtgcgtcc tctataagaa
      1261 tggagacag agtctatat ccagccacca caatgttctt gtcaggagtg cccagagtc
      1321 cccatttggt gtaatatgta cagaacctct gtggccctgt gggaaatccg aatcacccac
      1381 atgttgtctg tattcaagaa gcgactgagc aagacaaatt attttaactc aatgggtgct
      1441 ttataacttg agaacctctt ggtggcaca gatgagcatc tggcgctatc gctactatgc
30     1501 caggatttaa tatttgaatg actgtatttg ccaaagcaca taatacattt tgtttatcaa
      1561 atgtatttca ctataaaaaa aacaatctca tatacacgtg tacaatgact ggaacatctg
      1621 gagtacgttg cgggtgtgag tcctgttctg tgccatgttg tgtgtttgac caattaaatt
     1681 tattgaagag ataataaatc aatccctact tttatctgaa aaaaaaaaaa aaaaaaa

```

35 SEQ ID NO: 209

Amino acid sequence of mouse OPN3 encoded by the DNA sequence shown in SEQ ID NO: 208.

```

40     MYSGNRSGDQGYWEDGAGAEGAAPAGTRSPAPLFSPTAYERLALLLGCLALLGVGGNLLV
      LLLYSKFPRRLRTPHLFLVNLGLDGLVSLFGVTFTFASCLRNGVWDAVGCAWDGFSGS
      LFGFVSITTLTVLAYERYIRVVHARVINFSWAWRAITYIWLYSLAWAGAPLLGWNRYILD
      IHGLGCTVDWRSKDANDSSFVFLFLGCLVVPVGIIAHCYGHILYSVRMLRCVEDLQTIQ
      VIKMLRYEKKVAKMCFMAFVFLTCWMPYIVTRFLVNGYGHVLTPTVSVISYLFKSSST
      VYNPVIYIFMNRKFRSLLQLLCLRLRLRCQRPALDPAAGSEMHIRPIVMSQKDGDRPKK
      KVTFNSSSIIFIITSDESLSVDDSDKTNGSKVDVIQVRPL

```

45 SEQ ID NO: 210

ENSRNOT0000005072 cDNA sequence, EnsEMBL transcript [Rattus norvegicus]

```

      1 atgtactcgg ggaaccgtag cggcggccag ggctactggg aggacggggc gggcgccgag
     61 ggcgcagcac cggccggcac gcggagcccc gcgcctctct tcagcccccac cgcgtacgag

```

121 cgcctggcgc tgctgctcgg ttgcctcgcg ctgctgggcg tcggcggcaa cctgctggtg  
 181 ctgcttctct actccaagtt cccgcgactg cgcacgcccc cccacctctt cctgggtcaac  
 241 ctcagtcctgg gcgatctgct ggtatccctg ttcggagtca ccttcacctt cgcgtcgtgc  
 301 ctgcggaacg gctgggtgtg ggacgcctg ggctgcgctg gggacgggtt tagcggcagc  
 5 361 ctctttgggt ttgtttccat taccaccctc actgtgctgg cttatgaacg ttatatccgt  
 421 gtggtacatg ccagagtgat caattttttc tgggcctgga gggccattac ctatatctgg  
 481 ctctactctt tggcatgggc aggagcgcct ctctgggct ggaacaggta catcctcgat  
 541 gtacatggac tgggctgtac tgtggactgg aaatccaagg atgccaacga ctctctcttt  
 601 gtgctcttcc tgtttcttgg ctgtctgggt gtgcccattg gcatcatagc ccattgctac  
 10 661 ggtcatattc tgtattctgt tcgaatgctt cgctgtgttg aagatcttca gacaattcaa  
 721 gtgatcaaga tgttgcgata cgaaaagaaa gttgccaaga tgtgcttttt gatggccttt  
 781 ctctttctca cctgctggat gccttatggt gtaaccgcgt tcttggtggg caatggctat  
 841 ggacacctgg tcaccccaac tgtgtctatc gtttcttctc tctttgctaa atcgagcact  
 901 gtgtacaacc cagttatcta catcttcatg atcagaaaag ttcggagggt ccttctgcaa  
 15 961 ctctctgtgt tccgcctgct gagatgccag aggcctgcta aaaacctccc agcagctgag  
 1021 agtgaaatgc agatcaggcc catttgtgat tcacagaaaag atgggggacag gccaaagaa  
 1081 aaagtacatg ttaactcttc ctccatcatc ttcacatca ccagtgatga gtccctgtca  
 1141 gtcgaggaca gcgacagaag cagcgcctct aaggctgatg taatccaagt gcgtccttta  
 1201 tga

20 SEQ ID NO: 211

Amino acid sequence of rat OPN3 encoded by the DNA sequence shown in SEQ ID NO: 210.

MYSGNRSGGQGYWEDGAGAEGAAPAGTRSPAPLFSPTAYERLALLLGCLALLGVGNLLV  
 LLLYSKFPRLRTPHLFLVNLVSLGDLVSLFGVTFTFASCLRNGWVWDVGCWWDGFSGS  
 LFGFVSITTLTVLAYERYIRVVHARVINFSWAWRAITYIWLISLAWAGAPLLGWNRYILD  
 25 VHGLGCTVDWKSNDANDSSFVLFVFLGCLVPMGIIAHCYGHILYSVRMLRCVEDLQTIQ  
 VIKMLRYEKKVAKMCFMAFVFLTCWMPYVVTFRFLVNGYGHVLPVTVSIVSYLFAKSST  
 VYNPVIYIFMIRKFRRLQLLCFRLRCRPAKNLPAAESEMQIRPIVMSQKDGDRPKK  
 KVTFNSSSIIFIITSDESLVEDSDRSSASKVDVIQVRPL

SEQ ID NO: 212

30 gi|18860860|ref|NM\_006504.2| Homo sapiens protein tyrosine phosphatase, receptor type, E (PTPRE), transcript variant 1, mRNA

1 agccggagct ggagccgagg cggcggcggg acgcgccggg ccggacaaat ttctgtctag  
 61 gctgcggacg agcgggcggc aggagccggc gcgagcggct tcaggaaccc acggcctctg  
 121 cgcgtccccg cgaccttctc tcgcgccccg cgaagacagc cgggcgcccc ggagggcggc  
 181 gggcaggcgc ccgggagatg cggagcctcc gctgcagcgc gatctgcgcg accagaccgg  
 35 241 cccccccgag actatagcct tcactttccc tcggtccacc atggagccct tgtgtccact  
 301 cctgctggtg ggttttagct tgccgctcgc cagggtctctc aggggcaacg agaccactgc  
 361 cgacagcaac gagacaacca cgacctcagg cctccggac ccgggcgcct cccagccgct  
 421 gctggcctgg ctgctactgc cgtgctgctt cctcctctc gtgctcctc tcgccgccta  
 40 481 cttcttcagg ttcaggaagc agaggaaagc tgggttcagc accagcgaca agaagatgcc  
 541 caacggcaatc ttggaggagc aagagcagca aagggtgatg ctgctcagca ggtcacctc  
 601 agggcccaag aagtattttc ccatcccggt ggagcacctg gaggaggaga tccgtatcag  
 661 atccgcccag gactgcaagc agtttcggga ggagttcaac tcattgccat ctggacacat  
 721 acaaggaaact tttgaactgg caaataaaga agaaaacaga gaaaaaaca gatatccaa  
 45 781 catccttccc aatgaccatt ctagggtgat tctgagccaa ctggatggaa ttccctgttc  
 841 agactacatc aatgcttcc acatagatgg ttacaaagag aagaataaat tcatagcagc  
 901 tcaagggtccc aaacaggaaa cggtaaacga cttctggaga atggctctgg agcaaaagtc  
 961 tgcgaccatc gtcatgttaa caaacttgaa agaaaggaaa gaggaaaagt gccatcagta  
 1021 ctggcccgac caaggctgct ggacctatg aaacatccgg gtgtgcgtg aggatgcgt  
 50 1081 ggttttggtc gactacacca tccgggaagt ctgcatacag ccacagctcc ccgacggctg  
 1141 caaagccccc aggtggtct cacagctgca cttcaccagc tggcccgact tcggagtgcc  
 1201 ttttaccccc attgggatgc tgaagtctc caagaaagta aagacgctca acccctgca  
 1261 cgctgggccc atcgtggtcc actgtagcgc gggcgtgggc cggacgggca ccttcattgt

```

1321 gatcgatgcc atgatggcca tgatgcacgc ggagcagaag gtggatgtgt ttgaatttgt
1381 gtctcgaatc cgtaatcagc gccctcagat gggtcaaacg gatatgcagt acacgttcat
1441 ctaccaagcc ttactcgagt actacctcta cggggacaca gagctggacg tgcctccct
5 1501 ggagaagcac ctgcagacca tgcacggcac caccaccac ttcgacaaga tcgggctgga
1561 ggaggagtgc aggaattga caaatgtccg gatcatgaag gagaacatga ggacgggcaa
1621 cttgccggca aacatgaaga aggccagggt catccagatc atcccgtatg acttcaaccg
1681 agtgatcctt tccatgaaaa ggggtcaaga atacacagac tacatcaacg catccttcat
1741 agacggctac cgacagaagg actatttcat cgccaccag gggccactgg cacacacggg
1801 tgaggacttc tggaggatga tctgggaatg gaaatccac actatcgtga tgctgacgga
10 1861 ggtgcaggag agagagcagg ataaatgcta ccagtattgg ccaaccgagg gctcagttac
1921 tcatggagaa ataacgattg agataaagaa tgataccctt tcagaagcca tcagtatacg
1981 agactttctg gtcacttca atcagcccca ggcccgcag gaggagcagg tccgagtagt
2041 gcgccagttt cacttccacg gctggcctga gatcgggatt cccgccgagg gcaaaggcat
2101 gattgacctc atcgcagccg tgcagaagca gcagcagcag acaggcaacc accccatcac
15 2161 cgtgcactgc agtgccggag ctgggcgaac aggtacattc atagccctca gcaacatttt
2221 ggagcgagta aaagccgagg gactttttaga tgtatttcaa gctgtgaaga gtttacgact
2281 tcagagacca catatggtgc aaaccctgga acagtatgaa ttctgctaca aagtgggtaca
2341 agattttatt gatattttt ctgattatgc taatttcaaa tgaagattcc tgccttaaaa
2401 tattttttta tttaatggtc agtataattt gtaaaaatca tgtaattta tttcatagtt
20 2461 gacattaata tcttccctaa tttctttgt tatattttgt tatgccttaa aggccacctg
2521 ctatacagtt gttaaatctt aaatatgctt tttaaaaatt ggaataatgt attaaggtca
2581 aataatatcc cataaaatat atatttctgc taatattagt aaatatctta atttttaaaa
2641 aaaaaaaaaa aaaa

```

SEQ ID NO: 213

- 25 Amino acid sequence of human PTPRE encoded by the DNA sequence shown in SEQ ID NO: 212.

```

MEPLCPLLLVGFSPLALALRGNETTADSNETTTTSGPPDPGASQPLLAWLLLPLLLLLL
VLLLAAYFFRFRKQRKAVVSTSDKKMPNGILEEQEQQRVMLLSRSPSGPKKYFPIPV EHL
30 EEEIRIRSADDCKQFREEFNSLPSGHIQGT FELANKEENREKNRYPNILPNDHRSVILSQ
LDGIPCSYIDINASYIDGYKEKNK FIAAQGPKQETVNDFWRMVWEQKSATIVMLTNLKERK
EEKCHQYWPDQGCWYGNIRVCVEDCVVLVDYTIRKFCIQPLPDGCKAPRLVSQ LHTS
WPDFGVPTPIGMLKFLKKVKT LNPVHAGPIVHCSAGVGRGTGTFIVIDAMMAMMHA EQK
VDVFVFSRIRNRQRPQMVTDMQYTFIYQALLEYYLYGDTLVDSSLEKHLQTMHGT TTH
FDKIGLEEEFRKLTNVRIMKENMRTGNLPANMKKARVIQIIPYDFNRVILSMKRGQEYTD
35 YINASFIDGYRQKDYFIATQGPLAHTVEDFWRMIEWKSHTIVMLTEVQEREQDKCYQYW
PTEGSVTHGEITIEIKNDTLSEAISIRDFLVTLNQPPARQEEQVRVVRQFHFHGWPEIGI
PAEGKGMIDLIAAVQKQQQQTGNHPITVHCSAGAGRTGTFIALSNILERVKAEGLLDVFQ
AVKSLRLQRPHMVQTLEQYEFQYKVVQDFIDIFS DYANFK

```

SEQ ID NO: 214

- 40 gi|40805848|ref|NM\_130435.2| Homo sapiens protein tyrosine phosphatase, receptor type, E (PTPRE), transcript variant 2, mRNA

```

1 gtgcagcaga gggcagctga gaggtgggt ggctgggcct gggagacaca cagaggccag
61 gccttagcgc ggctcagcca tgagcaacag gactagcttt tcccggctca cctggttcag
121 gaagcagagg aaagctgtgg tcagcaccag cgacaagaag atgcccaacg gaactcttga
45 181 ggagcaagag cagcaaaggg tgatgctgct cagcagggtc ccctcagggc ccaagaagta
241 ttttcccatc cccgtggagc acctggagga ggagatccgt atcagatccg ccgacgactg
301 caagcagttt cgggaggagt tcaactcatt gccatctgga cacatacaag gaacttttga
361 actggcaaat aaagaagaaa acagagaaaa aaacagatat cccaacatcc ttccaatga
421 ccattctagg gtgattctga gccaactgga tgggaattccc tgttcagact acatcaatgc
50 481 ttcctacata gatggttaca aagagaagaa taaattcata gcagctcaag gtcccaaaaa
541 ggaacagggt aacgacttct ggagaatggt ctgggagcaa aagtctgcca ccatcgctcat
601 gttaacaaac ttgaaagaaa ggaaagagga aaagtgccat cagtactggc ccgaccaagg

```

661 ctgctggacc tatggaaaca tccgggtgtg cgtggaggac tgcgtgggtt tggctgacta  
 721 caccatccgg aagttctgca tacagccaca gctccccgac ggctgcaaaag cccccaggct  
 781 ggtctcacag ctgcacttca ccagctggcc cgacttcgga gtgcctttta cccccattgg  
 841 gatgctgaag ttcctcaaga aagtaaagac gctcaacccc gtgcacgctg ggcccatcgt  
 5 901 ggteccactgt agcgcgggcg tgggcccggac gggcaccttc attgtgatcg atgccatgat  
 961 ggccatgatg cagcgggagc agaaggtgga tgtgtttgaa tttgtgtctc gaatccgtaa  
 1021 tcagcgccct cagatggttc aaacggatat gcagtacacg ttcattctacc aagccttact  
 1081 cgagtactac ctctacgggg acacagagct ggacgtgtcc tccctggaga agcacctgca  
 1141 gaccatgcac ggcaccacca cccacttcga caagatcggg ctggaggagg agttcaggaa  
 1201 attgacaaat gtccggatca tgaaggagaa catgaggacg ggcaacttgc cggcaaacat  
 1261 gaagaaggcc aggggtcatcc agatcatccc gtatgacttc aaccgagtga tcctttccat  
 1321 gaaaaggggt caagaatata cagactatcc caacgcatcc ttcatagacg gctaccgaca  
 1381 gaaggactat ttcattcgcca cccaggggccc actggcacac acggttgagg acttctggag  
 1441 gatgatctgg gaattgaaat cccacactat cgtgatgctg acggagggtg aggagagaga  
 15 1501 gcaggataaa tgctaccagt attggccaac cgagggtcga gttactcatg gagaataaac  
 1561 gattgagata aagaatgata ccttttcaga agccatcagt atacgagact ttctggtcac  
 1621 tctcaatcag ccccaggccc gccaggagga gcagggtcga gtagtgccgc agtttcaactt  
 1681 ccacggctgg cctgagatcg ggattcccgc cgaggggcaaa ggcattgattg acctcatcgc  
 1741 agccgtgca aagcagcagc agcagacagg caaccacccc atcaccgtgc actgcagtgc  
 20 1801 cggagctggg cgaacaggta cattcatagc cctcagcaac attttgaggc gactaaagc  
 1861 cgagggaactt ttagatgtat ttcaagctgt gaagagttaa cgacttcaga gaccacatat  
 1921 ggtgcaaac ctggaacagt atgaattctg ctacaaagt gtacaagatt ttattgatat  
 1981 attttctgat tatgctaatt tcaaatgaag attcctgcct taaaatattt tttaatttaa  
 2041 tggctcagtat attttgtaaa aatcatgtta atttatttca tagttgacat taatatcttc  
 25 2101 cctaatttct ttgtatatat ttgttatgc cttaaaggcc acctgctata cagttggttaa  
 2161 atcttaataa tgctttttaa aaattggaat aatgtattaa ggtcaaataa tatcccataa  
 2221 aatatatatt tctgctaata ttagtaataa tcttaatttt tcattagatt catattctt  
 2281 aatttcacat attcaacacc tttaaatgtt gtaattctaa tatgcgaagt gtgcctctgc  
 2341 aagatactaa cacaagctc atgttaagaa aacagttgag gactcagaag tcagttgaaa  
 30 2401 atgcactttc ctaacagtga attcacaacc ctgaacagca gcatttttgg aaggcaaac  
 2461 gttcgtgatg gtacaatgta aatggggact tctgtaaaat tctcagtttc ggtccatgtg  
 2521 gtttatcttt acatttttaa gatcaaaaga gtcttttaca cctgaatcca ggtctaaaac  
 2581 acactagagt agctggtgac tataaataat attttaaaat gctgtgtcta caccatcaag  
 2641 actgtgtcta cactatcttg gctgaacgag aagagatgta aatgctgggt ggtcccgtt  
 35 2701 acccacggcg ttgggtacaa caaaaccagc catcgaggtt acaccccataa gcaccatttg  
 2761 ctgtccagct gcctgtcgtt tggcccagac caccctcaga aaaaaaccag ctgcctctcc  
 2821 cattctcccc tcccggtctg ccacagcggc ctgggctggg ccagtgtctat gcctggaggc  
 2881 tcaacacaaa acttcccac caaacattca gatgaactga gcgtcttaca cagcgagtac  
 2941 agaggagcac acattaggat agaaacagta gaataaccac gggcaattaa actttaaatt  
 40 3001 ttctgagcag cattttggta tttaaactt tcttgtaaaa agctgagaca gtttgtaaga  
 3061 aaagaatcct taaaatctag attttatcca ttttttaaaag tcccaccttt caatggttaa  
 3121 taaaacaaaa agagaaatcc ttaattctaa agctaaatta tttttgaaatg gaaatactac  
 3181 tgagaccatt gacactggat aacagtaatg atcccattac cagataagat tgactgacgg  
 3241 ggaaaaaaa aaaaaaagaa tggggtgtga atgtaccaac actgaatctt acagcagtta  
 45 3301 tctttctatg gccattaggt acctagcaga tgtgcacaat ataaacaaaa agatatctgg  
 3361 cctaccttac tactaaaagc atttaacacg tgcatttttg gtactttttt ttttggtttc  
 3421 taaaagctac ataaaggcct tatttgacat ttctactgat aactgatcgc acctcatgt  
 3481 tgcagtgttc gtcccttatt cattaacagt gtgtgcaaat gcaaccccag gacacattat  
 3541 tgttctgata gatgctcaca gggaatcaag cttctcgccc actccacggc tctgagcccc  
 50 3601 atccaagggc aagacttggg gccagctgg aaggacgaaa gcacactttg tgaccgccat  
 3661 cctcaccagc tgcattgcctg ggctgcacac tgcgtgaacg gctcctctct ccttctctgt  
 3721 acaatgattc agcatctcgg cgggaagagga aaatggagct ttttgaggct cgccagggtc  
 3781 cttttggttt caccattaaa attccaaacc caaagccttt gtttgactga aggagaagag  
 3841 aggggaagtaa gcttctgttc agcacctgaa ccctagaaaa agagccagtt tgctacgatg  
 55 3901 aagggtgacat ttctctggte atttatttga gagtctgaag tcaaagtcga ggggaccgg  
 3961 ctttggttca tgtctaggag ccctctgtca gaatccttga agccctttta agccttaact  
 4021 ggcattctct gtatcaagaa gtacctttaa ggtagacctt ttcagggtgc cctcaggaaa  
 4081 gggccctgct catgtttttt tctgtctccc ttagaccaac cccagggtgc cactgcaggg  
 4141 gttctgcctg ttcccaact ttttccattc caggacaaca ggagaagcca ctttcccag  
 60 4201 gacgcaagac tctccctcc actgtccggg acagcgttcg ccctttagcg gggagggtcat  
 4261 tacagcctca tggcctctac caaggcccca gatcacagga tctcctgggc cttggagcac

```

4321 ctcacgctgg gggaatcaat ccctgagggg ctcagaatct tctccgtgca acctggaaag
4381 ttcattctctt gtttccttca gtcaaagaaa gtccattgta cataacaaaa cagcccccaa
4441 acagcccagtg gccgacacca ttgttccttt cacactttcc ttgttgcat gcagttgggt
4501 tcaaattgcca aatagtgtt agaagacgac cattctgac tgtgtgtgat ctggctacta
5 4561 tgtgactgcc ttacgggtt ctctccatgt gctatatgaa tgaagaatgc ataccagtgt
4621 tttaaaagggt atttttatgt gtttttaaac acttttttaa atgagcctga cacctgtgtt
4681 tcagcatttg gagacatccc catgttatc ttttaagtgt ataattactg atactttttt
4741 gtttgtttgt ttaactaagt tgtgtttaac ttatgtgcag tctttataat gtatgtatgt
4801 tattacagtt tcaactatca tattttcttt gattacattt ataatttgat cttgctctga
10 4861 ttataatgcc agtgaatgtt gctgaactct ttgtatatgc aaattgcaag atttaaacca
4921 ttctgatgca aggataaacc ttactttga ctaccagcct gtgtttttgt ctttaaactc
4981 ctttaatttca ttctctgca aa

```

SEQ ID NO: 215

15 Amino acid sequence of human PTPRE variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 214.

```

MSNRSSFSRLTWFRKQKAVVSTSDKMPNGILEEQEQQRVMLLSRSPSGPKKYFPIPVE
HLEEEIRIRSADDCKQFREEFNSLPSGHIQGTPELANKEENREKNRYPNILPNDHSRVIL
SOLDGIPCSYIDYINASYIDGYKEKNKFLAAQGPKEQTVNDFWRMVWEQKSATIVMLTNLKE
20 RKEEKCHQYWPDPQGCWTYGNIRVCVEDCVVLVDYTIKFCIQPLPDGCKAPRLVSQLHF
TSWPDFGVFPFPIGMLKFLKKVKTILNPFVHAGPIVVHCSAGVGRGTGTFIVIDAMMAMHAE
QKVDVFVFSRIRNQRPMVQTMQYTFIYQALLEYYLYGDTLVDVSSLEKHLQTMHGT
THFDKIGLEBEFRKLTNVRIMKENMRGTGNLPANMKARVIQIIPYDFNRVILSMKRGQY
TDYINASFIDGYRQKDYFIATQGPLAHTVEDFWRMIWEKSHITVMLTEVQEREQDKCYQ
25 YWPTGVSHTGEITIEIKNDTLSEAISIRDFLVTNLNQPQARQEEQVRVVRQFHFHGWPEI
GIPAEGKGMIDLIAAVQKQQQTGNHPITVHCSAGAGRTGTFIALSNILERVKAEGLLDV
FQAVKSLRLQRPBMVQTLQYEFQYKVVQDFIDIFSDYANFK

```

SEQ ID NO: 216

gi|19171238|emb|AJ430580.1|HSA430580 Homo sapiens mRNA for tyrosine phosphatase epsilon PD1 (PTPRE gene)

```

30 1 ccttagcgcg gctcagccat gagcaacagg agtagctttt cccggctcac ctggttcagg
61 aagcagagga aagctgtggt cagcaccagc gacaagaaga tgcccaacgg aatcttgagg
121 gagcaagagc agcaaagggt gatgctgctc agcagggtcac cctcagggcc caagaagtat
181 tttcccatcc ccgtggagca cctggaggag gagatccgta tcagatccgc cgacgactgc
241 aagcagtttc gggaggagtt caactcattg ccatctggac acatacaagg aacttttgaa
35 301 ctggcaataa aagaagaaaa cacagataatc ccaacatcct tcccaatgac
361 cattctaggg tgattctgag ccaactggat ggaattccct gttcagacta catcaatgct
421 tcctacatag atggttataa agagaagaat aaattcatag cagctcaagg tcccaaacag
481 gaaacgggta acgacttctg gagaatggc tgggagcaaa agtctgcgac catcgctcatg
541 ttaacaaact tgaaagaaag gaaagaggaa aagtgccatc agtactggcc cgaccaaggc
40 601 tgctggacct atggaaacat ccgggtgtgc gtggaggact gcgtggtttt ggtcgactac
661 accatccgga agttctgcat acagccacag ctccccgacg gctgcaaagc ccccaggtcg
721 gtctcacagc tgcacttcac cagctggccc gacttcggag tgctttttac ccccatggg
781 atgctgaagt tcctcaagaa agtaaaagacg ctcaaccccg tgcaagctgg ccccatcggtg
841 gtccactgta gcgcgggctg gggccggagc ggcaccttca ttgtgatcga tgccatgatg
45 901 gccatgatgc acgcggagca gaagggtgat gtgtttgaat ttgtgtctcg aatccgtaat
961 cagcgccctc agatgggttca aacggatatg cagtacacgt tcatctacca agccttactc
1021 gagtactacc tctacgggga cacagagctg gacgtgtcct ccctggagaa gcacctgcag
1081 accatgcacg gcaccaccac ccacttcgac aagatcgggc tggaggagga gttcaggaaa
1141 ttgacaaatg tccggatcat gaaggagaac atgaggacgg gcaacttgcc ggcaaacatg
50 1201 aagaaggcca gggctcatcca gatcatccg tgtaaggcac ccgtggcggtg ccttggcgag
1261 ggctggggct gggctgggtgc cggaggcttt catcctggag aagccattga ccgcttaccc
1321 ctgtgcacca gggctcaaaag cagggtgtcc tctggcctca ggctaagggc tctgtgtgctg

```



```

1381 tgtaccaa at gtcaggtaga aaagacacct gtgaagcacc agctgaccca gacagtcctg
1441 catgggtctc aggcaagctg cttgggatat gagcttgtag atgcgggagg gacgccgagg
1501 ctgatttcag gggagggtct cacagaagct ggagatttgc ggcaggcacg acttgggatt
1561 cttgtctca ttccttgtct cccaagcct aggtgaactc agctgagacc acctaggaaa
5 1621 ggggcctgag ccacccagcc gagggccatgc aggtgaactc agctgagacc acctaggaaa
1681 aggtctgtggc cttgaggagg acaagaaatg ggcctagggt agagtcagga cactgggtctg
1741 ccacccaggt catgggagag ggaaggagga cctggcaggt gtgggaggag tggaggcaaa
1801 gctggcaaga agacagggcc acatgcagtg ggcaggggag tgggaatcgg gaggctcacc
1861 aagtgggaaa aagagaaaact tgagaaaccc ggcaccaaag cagactccac agggatccga
10 1921 ggaactaggg agggagggtg gagatgcccc gcttgggagg aggggtgtgccc tcgttataga
1981 cttgtcctcg acagagtcac agctcaacaa acaccatctg gaaacagctg ggagcaggcc
2041 tgctgcagca ggggtctcccc gagcactgca gatgacttca accgagtgat cctttccatg
2101 aaaaggggtc aagaatacac agactacatc aacgcacatc tcatagacgg ctaccgacag
2161 aaggactatt tcatcgccac ccaggggcca ctggcacaca cgggtgagga cttctggagg
15 2221 atgatctggg aatggaagtc ccacactatc gtgatgctga cggagggtga ggagagagag
2281 caggataaat gctaccagta ttggccaacc gaggggtcag ttactcatgg agaaataacg
2341 attgagataa agaatagata cctttcagaa gccatcagta tacgagactt tctggtcact
2401 ctcaatcagc cccaggcccc ccaggaggag caggtccgag tagtgcgcca gtttacttc
2461 cacggctggc ctgagatcgg gattcccgcg gagggcaaaag gcatgattga cctcatcgca
20 2521 gccgtgcaga agcagcagca gcagacaggg aaccacccca tcaccgtgca ctgacgtgcc
2581 ggagctgggc gaacaggtac attcatagcc ctcagcaaca ttttggagcg agtaaaagcc
2641 gagggacttt tagatgtatt tcaagctgtg aagagtttac gacttcagag accacatatg
2701 gtgcaaaccc tggaaacagta tgaattctgc tacaagtggt tacaagattt tattgatata
2761 ttttctgatt atgctaattt caaatgaaga ttctgcctt aaaatatttt ttaatttaat
25 2821 ggtcagtata ttttgtaaaa atcatgttaa tttatttcat agttgacatt aatatcttcc
2881 ctaatttctt tgtatatatt ttgttatgcc ttaaaggcca cctgctatac agttgttaaa
2941 tcttaaatat gctttttaaa aattggaata atgtattaag gtcaaataat atcccataaa
3001 atatatattt ctgctaataa aaaaaaactt t

```

SEQ ID NO: 217

- 30 Amino acid sequence of human PTPRE variant ORF number 2 encoded by the DNA sequence shown in SEQ ID NO: 216.

```

MSNRSSFSRLTWFRKQRKAVVSTSDKMPNGILEEQEQQRVMLLSRSPSGPKKYFPPIPVE
HLEEEIRIRSADDCKQFREEFNSLPSGHIQGT FELANKBENREKNRYPNILPNDHSRVIL
SOLDGIPICSDYINASYIDGYKEKNKFLAAQGPKEQTVNDFWRMVWEQKSATIVMLTNLKE
35 RKEEKCHQYWPDQGCWTYGNIRVCVEDCVVLVDYTIKFCIQQLPDGCKAPRLVSQLHF
TSWPDFGVFPFTPIGMLKFLKKVKTLNPVHAGPIVHCSAGVGRGTGTFIVIDAMMAMMHA
QKVDVFEEFVSIRNRQRPQMVTDMQYTFIYQALLBYLYGDTLDSVLSLEKHLQTMHGT
THFDKIGLEEFKRLTNVRIMKENMRTGNLPANMKKARVIQIIPCKAPVAWLGGQGWAG
AGGFHPGEAIDRLPLCTRVKSRVSSGLRLRAPVCCTKCQVEKTPVKHQLTQTVLHGSQAS
40 CLGYELADAGGTPRLISGEGFTEAGDLRQARLGIPCLIPCLPKPGQRPASPLHRGA

```

SEQ ID NO: 218

gi|199932|dbj|D83484.1|MUSPTPE Mouse mRNA for protein tyrosine phosphatase epsilon, complete cds

```

1 1 caccgggtccc agccctgctt taccagttgc gaaaccttaa gtgccacact actttgtggt
45 61 gtncagagct aaggctaaat acgaaggagc gttgagtttt ctctatggct cgtctgagtg
121 ccatcacctt ttaggaccc cctgataga gtgtcagtc gagtgagga gagaggcatg
181 ttctgcac cctccagtg aaagccaaag ctccagaggg tgagcgccgc gtgcgctccc
241 agagccaggc accaggcgtc gccgtcgccc ctgcagctc caccggccccg ccccnccgc
301 ggagctggag ctggagccgc agccggagcc ctagcccaga gctggaggcg gccggaccgg
50 361 gccagacaga ttcctgctg tcctgtagg ctgagggtg cgggtgcgg gctacgggct
421 gcgggctaca ggctacggc tacggactgc gggctgcgag agaagtcacc ggcggcgag
481 gtacctcact gcctacgcgc gtccccacga ccctccctc cgccggcgagg gacagcggag

```

541 gccccggagc ggagcgcgct gggcgggcgc ccgggagatg cggagtcgcc gcggcggagc  
 601 gatcggggct acagcaccgg tccctgggag actatagccc tcatctcccc ttggtgcacc  
 661 atggagccct tctgtccact cctgctggca agtttttagct tgctcgctcg cagagctggc  
 721 cagggaacag acaccacccc aacagagagc aactggacca gcacaactgc aggccctccg  
 5 781 gacctgtgtg catcccagcc gctgctcacc tggctgctgc tgccctgct cctcctcctg  
 841 ttctgtcttg cagcctactt ctccaggttc cggagcgaga ggaaggccgt ggtcagcagc  
 901 aacgacaaga aaatgcctaa cgggatctta gaagagcaag agcagcagag agtgatgctg  
 961 ctgagcagat ctccatcagg cccaagaag ttcttcccca tccccgtgga gcacctggag  
 10 1021 gaggagatcc ggggtgagatc tgcggatgac tgcaagcgat tccgagagga gttcaattca  
 1081 ttgccatctg gacacataca aggaaccttt gaactagcaa ataaagaaga aaacagagaa  
 1141 aaaaacagat accccaacat tctgccaat gatcattgca gagtgatttt gagccaagtg  
 1201 gatggaatcc cctgctctga ctgacctaat gcttcctaca tcgatggcta caaagaaaag  
 1261 aacaaattca tagcagctca aggccctaag caggagacag tgaatgactt ctggagaatg  
 1321 gtctgggagc aaaggctcagc caccatcgct atgttgacga acctgaagga gaggaaggag  
 15 1381 gagaagtgtc accagtactg gccagaccag ggctgttggc cctacggcaa catccgggtg  
 1441 tgtgtagagg actgctgtgt cctggtggat tacacgatcc gaaagtcttg catccatccg  
 1501 caactcccag acagctgcaa agcccccgcg ctggtctcac agctgcactt caccagctgg  
 1561 cctgacttcc ggggtgccgtt tacccccact gggatgctca agttcctgaa gaaagtgaag  
 1621 acatcaacc cctcacatgc tgggccatt gtggttcact gtacgcggg cgtgggtcgg  
 20 1681 actggcacct tcattgtgat cgatgccatg atggacatga tacactcggg cagaagggtt  
 1741 gacgtctttg agtttgtgtc tagaatccgc aatcagcgcc ctcatatggt ccagacggat  
 1801 gttcagtata cattcatcta ccaagcctta ctggaatact acctctatgg ggacacagag  
 1861 ctggatgtgt cctccctgga gaggcacctg cagacgctcc atagcacagc caccattttt  
 1921 gacaagatcg ggctggagga agagttcagg aagctgacca acgtgcgaat catgaaggag  
 25 1981 aacatgagga cgggcaacct gcctgccaac atgaagaagg cccgcgtcat ccagatcatt  
 2041 ccatatgact tcaatcgggt catcctgtcc atgaaaagag ggcaagagtt cacagactat  
 2101 atcaacgcat ccttcataga tggctacagg cagaaggact acttcatggc cacacaggcg  
 2161 cctctggctc acacagttga ggacttctgg aggatgggtat gggagtggaa gtctcacaca  
 2221 atgctcatgc tgacggaggt gcaggagcgg gaacaggata aatgctacca gtattggcca  
 30 2281 acggagggct cggtgactca tggagatata actatagaga taaagagcga caccctgtct  
 2341 gaagcaatca gcgtacgaga ctttctgggt acctcaaac agccccctggc ccgccaggaa  
 2401 gagcaggtcc gcatggtgag acaattccat ttccatggct ggctgaggt tggcatcccc  
 2461 gctgaaggca aaggcatgat tgacctgat gcagcagtg agaaagcagc gcagcagc  
 2521 ggcaaccacc ccatcacctg gacctgcagc gcgggagcag ggcggaagcgc tacattcata  
 35 2581 gcactcagta acatttttga acgagtgaac gccgagggac tcctagacgt gtttcaagct  
 2641 gtgaagagct taagacttca gagaccacac atggtgcaga ccctggagca atatgaattc  
 2701 tgctacaaag tgggtacaaga ttttatcgat atattttctg attatgctaa tttcaaatga  
 2761 agattcctgc cttaaaatat tttttaattt aatggctcag atatttttga aaaaaaaaaa  
 2821 aaaaaaa

40 SEQ ID NO: 219

Amino acid sequence of mouse PTPRE encoded by the DNA sequence shown in SEQ ID NO: 218.

45 MEPFCPLLLASFSLSLARAGQGNDTTPTESNWTSTTAGPPDPGASQPLLWLLPLLLLL  
 FLLAAYFFRFRKQKAVVSSNDKKMPNGILEEQEQQRVMLLSRSPSGPKKFFPIVEHLE  
 EEIRVSADDCRFRFEEFNSLPSGHIQGTPELANKEENREKNRYPNILPNDHCRVILSQV  
 DGIPCSYDINASYIDGYKEKNKFIAAQGPQKETVNDFWRMVWEQRSATIVMLTNLKERKE  
 EKCYQYWPDQGCWYGNIRVCVEDCVVLVDYTIKFCIHPQLPDSCKAPRLVSQLHFTSW  
 PDFGVFFTFIGMLKFLKKVKTLPNSHAGPIVVHCSAGVGRGTGFIVIDAMMDMIHSEQKV  
 50 DVFEFVSRIQRNPQMVQTDVQYTFIYQALLEYYLYGDTLQVSSLERHLQTLHSTATHF  
 DKIGLEEFRLKLTNVRIMKENMRTGNLPANMKARVIQIIPYDFNRVILSMKRGQEFQDY  
 INASFIDGYRQKDYFMATQAPLAHTVEDFWRMVWEKSHTMLMLTEVQEREQDKCYQYWP  
 TEGSVTHGDITIEIKSDTLSEAI SVRDFLVTFKQPLARQEEQVRMVRQFHFHGWPEVGIP  
 AEGKGMIDLIAAVQKQQQTGNHPITVHCSAGAGRTGTFIALSNILERVKAEGLLDVFAQ  
 VKSLRLQRPHMVQTLQYEFQYKVVQDFIDIFSDYANFK

55 SEQ ID NO: 220

gi|34861150|ref|XM\_341950.1| Rattus norvegicus Protein tyrosine phosphatase, receptor type, epsilon polypeptide (Ptpre), mRNA

```

1 atggaaccct tctgtccact cctgttgga agtttttagct tgcgcttgc cacagctggc
5 61 cagggcaacg acaccacccc aacagagagc aactggacca gtacaactgc aggtcctccg
121 gaccctggca cgtcccagcc gctgcttacc tggctgctgc tgccctctgt cctcctcctg
181 ttcctccttg cagcctactt cttcaggttc aggaagcaga ggaaagccgt ggtcaacagc
241 aacgacaaga agatgcctaa cgggatccta gaagaacaag agcagcagag agtgatgcta
301 ctgagcagat ctccatcggg tcccaagaag tacttcccca tccctgtgga gcacctggag
361 gaggagatcc ggggtgagatc tgcggatgac tgcaagcggg tccgagagga gttcaattca
10 421 ttgccatccg gacacatata aggaaccttt gaactagcaa ataaagaaga aaacagagaa
481 aaaaacagat accccaacat tctgcccatt gatcattgca gaggatgtt gagccaattg
541 gatggaatcc cctgctctga ctacattaat gcttcctaca tagatggcta caaagaaaag
601 aacaaattca tagcagctca agggcccaag caggagacag tgaatgactt ctggagaatg
661 gctctggcgc aaaggtcagc caccattgtc atgttgacga acctgaagga aaggaaggag
15 721 gagaagtgtt accagtactg gccagaccag ggctgttgga cttatgggaa catccgggtg
781 tgtgtagagg actgctgtgt cctggtggat tacaccatcc gcaagtctg catccatccg
841 caactccccg acagctgcaa agctccacgg ctggtctcac agctgcactt caccagctgg
901 cctgactttg gggtaccatt taccaccaatc ggaatgctga agttcctgaa gaaagtgaag
961 aactcaacc cctcacatgc tgggccatt gtggtccact gtagcgcggg cgtgggtcgg
20 1021 actggcacct tcattgtgat cgatgccatg atggacatga tccactcgga acagaaggtt
1081 gacgtcttcg agtttgtgtc tagaattccg aatcagcgcc ctcataggtt ccagacagat
1141 gttcagcata cattcatcta ccaagcctta ctggaatact acctctatgg ggacacggag
1201 ctggatgtgt cctccttggg gaggcacctg cagacgctac atggcacagc caccattttt
1261 gacaagatcg ggctggagga agagttcagg aaactgacca acgtgcgaat catgaaggag
25 1321 aacatgagga cgggcaacct gccggccaac atgaagaagg cccgtgtaat ccagatcatc
1381 ccatatgact tcaatcgggt aatcctgtcc atgaaaagag ggcaagagtt cacagactac
1441 atcaatgcat ccttcataga tggctacagg cagaaggact acttcatggc tacacagggg
1501 cctctggcgc acacagttga ggacttctgt aggatgggtc gggagtggaa gtcacacaca
1561 atcgtcatgc tgactgaggt gcaggagcgg gagcaggata aatgctacca gtattggcca
30 1621 accgagggct cagtgactca cggagatata actatagaaa taaagagtga caccctgtct
1681 gaagcaatca gcatacgaga ctttctggtt actttcaaac agccccctggc ccgccaggaa
1741 gagcaggtcc ggatgggtgag acaattccat ttccatggct ggctgaggt tggaaatcccc
1801 actgagggta aaggcatgat tgacctgac gcggcagtg agaaagcagc gcagcagaca
1861 ggcaaccacc ccatcacctg gcactgcagt gcgggagcag ggcgagacag tacattcata
35 1921 gcactcagta acatttttga acgagtgaag gccgagggag tcctagatgt gtttcaagct
1981 gtgaagagct taagacttca gagaccacac atgggtgcaa ccctggagca atatgaattc
2041 tgctacaaag tgggtacaaga ttttatcgat atattttctg attatgctaa tttcaaatga
2101 agatccctgc cttaaaatat tttttaattt aatggctcag atatttttga aaaaaaatc
2161 atgttaattt atttcctagt ggatattaat atccgtcctg attcctttgt atatatttt
40 2221 ttatgttcta aaggctacct gctgaagat tattaatat taaatgtcct tttt

```

SEQ ID NO: 221

Amino acid sequence of rat PTPRE encoded by the DNA sequence shown in SEQ ID NO: 220.

```

45 MEPPCPLLLASFSLATAGQGNDTTPTESNWTSTTAGPPDPGTSQPLLTWLLPLLLLL
FLLAAYFFFRKQKAVVNSNDKKMPNGILEEQEQQRVMLLSRSPSGPKKYFPIPVEHLE
EEIRVRSADDCKRFREFNSLPSGHIQGTPELANKEENREKNRYPNILPNDHCRVILSQL
50 DGIPCSDYINASYIDGYKEKNKFIQAQGPQKETVNDPFRMVWVQRSATIVMLTNLKERKE
EKCYQYWPDPQGCWYTGNIIRVCVEDCVLVDTYIRKFCIHPQLPDSCKAPRLVSQLHFTSW
PDFGVPTPIGMLKFLKKVKTLPNSHAGPIVVHCSAGVGRGTGTFIVIDAMMDMIHSEQKV
DVPEFVSRIQRNRPQMVQTDVQYTFIYQALLEYYLYGDELVDVSSLERHLQTLHGTHF
DKIGLEBEFRKLTNVRIMKENMRTGNLPANMKKARVIQIIPYDFNRVILSMKRGQEFTDY
INASFIDGYRQKDYFMATQGPLAHTVEDPFRMVWVWKSHTIVMLTEVQEREQDKCYQYWP
TEGSVTHGDITIEIKSDTLSEAISIRDPLVTFKQPLARQEEQVRMVRQFHFHGWPEVGIP
55 TEKGGMIDLIAAVQKQQQTGNHPITVHCSAGAGRTGTFIALSNI LERVKAEGLLDVFA

```

VKSLRLQRPHMVQTLQYEFQYKVVQDFIDIFSDYANFK

SEQ ID NO: 222

gi|31083343|ref|NM\_020311.1| Homo sapiens G protein-coupled receptor (RDC1), mRNA

```
1  tgcaagtctg  cagccagcag  agctcacagt  tggtgcaaag  tgctcagcac  taagggagcc
5  61  agcgcacagc  acagccagga  aggcgagcga  gcccgagccag  cccagccagc  ccagccagcc
121  cggaggtcat  ttgattgccc  gcctcagaac  gatggatctg  catctcttcg  actactcaga
181  gccagggaac  ttctcggaca  tcagctggcc  atgcaacagc  agcgactgca  tcgtgggtgga
241  cacggtgatg  tgtcccaaca  tgcccaacaa  aagcgctctg  ctctacacgc  tctccttcat
301  ttacattttc  atcttcgtca  tcggcatgat  tgccaactcc  gtggtgggtc  ccatgcccga
10  361  ccaggccaag  accacaggct  atgacacgca  ctgctacatc  ttgaacctgg  ccattggccga
421  cctgtggggt  gtccctacca  tcccagctcg  ggtggtcagt  ctgctgcagc  acaaccagtg
481  gcccatgggc  gagctcacgt  gcaaagtcac  acacctcacc  ttctccatca  acctcttcgg
541  cagcattttc  ttctctacgt  gcatgagcgt  ggaccgctac  ctctccatca  cctacttcac
601  caacaccccc  agcagcagga  agaagatggt  acgcccgtgc  gtctgcaccc  tggtgtggct
15  661  gctggccttc  tgcgtgtctc  tgcctgacac  ctactacctg  aagaccgtca  cgtctgcgtc
721  caacaattag  acctactgcc  ggtccttcta  ccccgagcac  agcatcaagg  agtggctgat
781  cggcatggag  ctgggtctcc  ttgtcttggg  ctttgccgtt  ccttctccca  ttatcgctgt
841  cttctacttc  ctgctggcca  gagccatctc  ggcgtccagt  gaccaggaga  agcacagcag
901  ccggaagatc  atcttctcct  acgtgggtgt  cttccttgtc  tgctggctgc  cctaccacgt
20  961  ggcggtgctg  ctggacatct  tctccatcct  gcactacatc  cctttcacct  gccggctgga
1021  gcacgccttc  ttcacggccc  tgcattgtac  acagtgcctg  tcgctgggtc  actgctgcgt
1081  caaccctgtc  ctctacagct  tcatcaatcg  caactacagg  tacgagctga  tgaaggcctt
1141  catcttcaag  tactcggcca  aaacagggtc  caccaagctc  atcgatgcct  ccagagtctc
1201  agagacggag  tactctgcct  tggagcagag  caccaaatga  tctgccctgg  agaggctctg
25  1261  ggacgggttt  acttggtttt  gaacagggtg  atgggccccta  tggttttcta  gagcaaagca
1321  aagtagcttc  ggggtcttgat  gcttgagttag  agtgaagagg  ggagcacgtg  cccctgcat
1381  ccattctctc  tttctcttga  tgacgcagct  gtcatttggc  tgtgcgtgct  gacagttttg
1441  caacaggcag  agctgtgtcg  cacagcagtg  ctgtgcgtca  gagccagctg  aggacaggct
1501  tgcttggact  tctgtaagat  aggattttct  gtgtttcctg  aattttttat  atggtgattt
30  1561  gtattttaa  ttttaagact  tattttctca  ctattggtgt  acctataaaa  tgtatttgaa
1621  agttaaatat  attttaata  ttgtttggga  ggcatagtgc  tgacataat  tcagagtgtt
1681  gtagttttta  ggtagcgtg  acttcagttt  tgactaagga  tgacactaat  tgttagctgt
1741  tttgaaatta  tatatatata  aatatataaa  aatatataaa  tatatgccag  tcttggtgta
1801  aatgttttat  ttaccatagt  tttatatctg  tgtggtgttt  tgtaccggca  cgggatatgg
35  1861  aacgaaaact  gctttgtaat  gcagttttgt  acattaatag  tattgtaaag  ttacatttta
1921  aaataaacia  aaaactgttc  tggactgcaa  atctgcacac  acaacgaaca  gttgcatttc
1981  agagaattct  ctcaatttgt  aagttatttt  tttttaataa  agatttttgt  ttccaaaaaa
2041  aaaaaaaaaa  aaaaaaaaaa  aaaa
```

SEQ ID NO: 223

40 Amino acid sequence of human RDC1 encoded by the DNA sequence shown in SEQ ID NO: 222.

```
MDLHLFDYSEPGNFSDISWPCNSSDCIVVDVMCPNMPNKSULLYTLFSFIYIFIVIGMI
ANSVVVVVNIQAKTTGYDTHCYILNLAIADLWVLTIPVWVSVLVQHNQWPMGELTCKVT
HLIFSINLFGSIFFLTCMSVDRLSITYFTNTPSSRKKMVRVVVCILVWLLAFVSLPDT
45  YYLKTVTASNNETYCRSFYPEHSIKEWLIGMELVSVVLGFVFPFSIIAVFYFLARAI
ASSDQEKHSSRKIIFSYYVVFVLCWLPYHVAVLLDIFSILHYIPFTCRLEHALFTALHVT
QCLSLVHCCVNPVLYSFINRNYRYELMKAFIFKYSAKTGLTKLIDASRVSTETYSALEQS
TK
```

SEQ ID NO: 224

gi|31560714|ref|NM\_007722.2| Mus musculus chemokine orphan receptor 1 (Cmkr1), mRNA

```

1  gctcagttgc tacaaactgc tcagcactga aggagcctgc agcgctcacc gtcaggaagg
61  caaaccacag cccaggaagc cctgagggtca cttgggtcgct ctccctcaaga ccatggatgt
5  121  gcacttgttt gactatgcag agcctgggcaa ctactctgac atcaactggc catgtaacag
181  cagcgactgc attgtgggtg acactgtgca gtgtcccacc atgcctaaca agaacgtgct
241  tctgtatacc ctctccttca tctacatttt catcttcgtg atcggcatga ttgccaacte
301  tgtgggtggc tgggtgaata tccaggctaa gaccacaggc tacgacacgc actgctacat
361  cttgaacctg gccattgcag acctgtgggt cgtcatcacc atccccgtct ggggtggtcag
10  421  tctcgtgcag cataaccagt ggcccatggg ggagctcaca tgcaagatca cacacctcat
481  tttctccatc aacctctttg ggagcatctt ctctctcgcc tgcagagcg tggaccgcta
541  tctctccatc acctacttca ccggcacctc cagctataag aagaagatgg tacgccgtgt
601  tgtatgcac ttgggtgtggc tgctggcctt ctttgtgtcc ctgcctgata cctactacct
661  gaagacgggc acatctgctt ccaacaatga gacctactgc aggtccttct accccgagca
15  721  cagcatcaag gagtgggtga tccggcatgga gctgggtctct gtcatcttgg gcttctgtgt
781  ccccttccat atcattgcga tcttctactt cctgctcgct agagccatgt cagcatcagg
841  cgaccaggag aagcacagta gccggaagat catcttctcc tacgtgggtg tcttctcggt
901  atgttggtcg ccgtaccatt ttgtggttct gctggacatc ttctccatct tacactacat
961  cccgtttacc tgtcagctgg agaattgtgt ctttacagcg ttgcatgtca cccagtgcct
20  1021  gtctctgggt cactgctgtg tcaaccccg tctctacagc ttcacacacc gcaactacag
1081  gtacgagctg atgaaggcct tcatcttcaa gtactcggcc aaaacaggtc tcaccaagct
1141  cattgatgcc tccagagtgt cagagacaga gtactctgcc ctggaacaga acaccaagtg
1201  atccatcatt ctgcagaggt cggggggaca cgtgcatggt gcaaatgggg cggctgggtc
1261  ctgtggtttc ttcaagaaag caatgtagct ttgggtctgg ttgcttgagt ggtatgaaga
25  1321  ggagaaggca tgtgaccatg ctctctgtct ctcatctccc agccaagatg gctgccgcct
1381  ggatgcccac cctgacagct ggcagctggc agctggcagc tggcagcagg cagggtgtgt
1441  ctgtgctgtg ctgtgctgct ctgtgctgtg ctgtgcagcc agagctgcgt atcaaagcca
1501  gcaccaggac aggtctctat ggacatgtgt acagtagaat cttctgtgtt tcttcaagtt
1561  tttacttggg gactttttgt atttaagttt taagacttta tttctcact atggatgtac
30  1621  cttataaatg tatttgaaag ctaaataat tttaaatatt gtatgggagg tgaaggctg
1681  ttgtattcag accatgtagg cctcagatta gctggacttg agttttgact aaggatgaca
1741  ttaattgtta gctgatttga aattatatat ataaatataa aaatatatat aaatttatgc
1801  cagtctcggc tgaaagggtt tatttacaat agttttatat ctgtgcaatg ttaatgctg
1861  gcacagcata tgaaacaaca actgccagc aatgcagttt gttacatgaa ccatattgta
35  1921  gagttacact tcgggtgcga aaccgctgaa caagggaac agagcggtct gttgatttgt
1981  aagttatttt ttaataaag attttgttct ctaaaacctc atgtgtgctt tgggtttgaa
2041  agcatgtttg ggctttggat ggcattgcata gactaacac atggaaaaac atgtcagaca
2101  caggcaagaa gaagatgttc agtgggagag aacggccaga gagcccaacc atgggactgg
2161  gtggatggat ctgagcaggg tgggtgctat tccttggttc ctgaaggaac attttcaaaa
40  2221  acatttgagt tcaggggagg atgagcaggt cagaggtagg tgcatttcaa tggctgttga
2281  gtgaaacttg ctgtaaaagt tgtgctgccc actgtctact caggaaatca tgccttatac
2341  ttaacaaaca tctgcctgcc gtcttcccca gccacatgtc catcagacca tcctgtgag
2401  ggcctgaggt taggtgtcag cgagctgtgt atgctgtatg actctccaag cccaaagcac
2461  cttcaggccc agcactctag agtcagaggg aggtggatct ctgtgaggtt gaggccaggg
45  2521  ccaacctgat ctacacagtg agttccaggg tagccaggcc tgtgtagtaa gtctgtgtct
2581  g

```

SEQ ID NO: 225

Amino acid sequence of mouse RDC1 encoded by the DNA sequence shown in SEQ ID NO: 224.

```

50  MDVHLFDYAEFGNYSDINWPCNSSDCIVVDTVQCPTMPNKNVLLYTLFSFIYIFIVIGMI
ANSVVVWVNIQAKTTGYDTHCYILNLAIDLWVVITIPVWVVS LVQHNQWPMGELTCKIT
HLIFSINLFGSIFFLACMSVDRYLSITYFTGTSSYKKMVR RVVCILVWLLAFFVSLPDT
YYLKT VTSASNNETYCRSFYPEHSIKEWLIGMELVSILGF VAVPFTIIAIFYLLARAMS
ASGDQEKHSSRKIIFS YVVVFLVCWLPYHFVLLDIFSILHYIPFTCQLE NVLFTALHVT
55  QCLSLVHCCVNPVLYSFINRNYRYELMKAFIFKYSAKTGLTKLIDASRVSETEYSAL EQN

```

TK

SEQ ID NO: 226

gi|16758073|ref|NM\_053352.1| Rattus norvegicus chemokine orphan receptor 1 (Rdc1), mRNA

```

5      1 cactgaagga gcctgcagca ctcaaagcca ggaaggcaaa tcacagccca ggcagccgcg
      61 aggtcacttg gttgctctcc tcaagaccat ggatgtgcat ctgtttgact atgtggaacc
     121 tgggaactac tcggacatca actggccctg taacagtagc gactgcatcg tcgtggacac
     181 cgtgcagtgt cccgccatgc ccaacaagaa tgtgctgctg tataccctct ccttcatacta
     241 cattttcatc ttcgtgatcg gtatgattgc caactccgtg gtggtctggg tgaatatcca
10    301 ggccaagact acaggctacg acacacactg ctacatcttg aacctggcca ttgctgatct
     361 gtgggtcgtc atcaccatcc ctgtctgggt ggtcagtcct gtgcagcata accagtggcc
     421 catgggtgag ctacagtgca agatcacaca cctcattttc tccatcaacc tctttgggag
     481 catcttcttc ctcgcatgca tgagcgtgga ccgctatctc tccatcacct acttcaccag
     541 cacctccagc tataagaaga agatgggtacg ccgtgttgtc tgcgtcttgg tgtggctgct
15    601 ggccttcttt gtgtccctgc ctgacaccta ctacctgaag acggtcacat ctgcttccaa
     661 caacgagacc tactgcaggt ccttctaccc cgagcacagc atcaaggagt ggctcattgg
     721 catggagctg gtctccgtca tcttgggttt tgctgtcccc ttcaccatca ttgctatctt
     781 ctacttctcg ctgccagag ccatgtcagc atccggtgac caggagaaac acagcagccg
     841 gaagatcatc ttctcctacg tgggtggtctt cctgggtgtg tggctgccgt accattttgt
20    901 ggttctgctg gacatcttct ctatcttgca ctacatcccg ttcacctgcc aactggagaa
     961 tgtgctcttt acagcgtgc acgtcacgca gtgcctgtcc ctggtgcaact gctgtgtcaa
    1021 ccctgtgctc tacagcttca tcaaccgaaa ctacaggtag gagctgatga aggccttcat
    1081 cttcaagttac tcagccaaaa caggactcac caaactcatc gatgcctcca gagtgtcaga
    1141 gacagagtac tctgccctgg agcagaacac caagtgaccg tgctatagag gcattgggac
25    1201 atgtgcatgt tgcaaatggg gcagctgggc cctgtgggtt cttcaagaaa gcactgtagc
    1261 tttgggtctg gttgcttgag tggatgaag aggagaagga ccatgcttcc tgtctctcat
    1321 tctccagcca agatggctgt cacctggaca ctcatcctga cagcttgca cagacaggct
    1381 gccctgcgct gtgcagccag agctgtgtgt caaagccagc atcaggacag actcttctgg
    1441 acacgtgtac aatagaatct tttgtgtttc ttcaagtttt tacttggtga cttttgtatt
30    1501 taagtthttaa gactttatth tctcactatg gatgtacct ataaatgcat ttgaaagcta
    1561 aatatatthtt aaatatthtt tgggaggtgt aaggctgttg tatttgacc atgtaggcct
    1621 cagattagca ggactctgag ttttgactaa ggatgacatt cattgttagc tgattcgaaa
    1681 tgatataaat atatatataa tataaatata tatataaatt tatgccagtc ctggctggaa
    1741 tgtttttatth acaatagthtt tatatctgtg tgggtgtthaa tgctggcaca gcatatggaa
35    1801 cggaaactgc ccagcaatgc agtttgtgac acgaaccgta tgtagagtta cgtttcgggc
    1861 tgcaaacccg tgaacagagg aaacagcatt ctcttgattg taagttatth ttttaataaa
    1921 gatttttgtt tcctaaaaaa aaaaaaaaaa aaaaaaaa

```

SEQ ID NO: 227

Amino acid sequence of rat RDC1 encoded by the DNA sequence shown in SEQ ID NO: 226.

```

40 MDVHLFDYVEPGNYSDINWPCNSSDCIVVDTVQCPAMPNKNVLLYTLFSFIYIFVIGMI
   ANSVVVVWNIQAKTTGYDTHCYILNLAIDLWVVITIPVWVVS LVQHNQWPMGELTCKIT
   HLIFSINLFGSIFFLACMSVDRYLSITYFTSTSSYKKKMVRRVVCVLVLLAFFVSLPDT
   YYLKTVTSASNNETYCRSFYPEHSIKEWLIGMELVSVILGFVAVPFTIIAIFYLLARAMS
   ASGDQEKHSSRKIIIPSYVVVFLVCWLPYHFVLLDIFSILHYIPFTCQLENVLF TALHVT
45 QCLSLVHCCVNPVLYSFINRNYRYELMKAFIFKYSAKTGLTKLIDASRVSETEYSALEQN
   TK

```

SEQ ID NO: 228

gi|4759145|ref|NM\_004787.1| Homo sapiens slit homolog 2 (Drosophila) (SLIT2), mRNA

1 cagagcaggg tggagagggc ggtgggagggc gtgtgcctga gtgggctcta ctgccttgtt  
 61 ccatattatt ttgtgcacat tttccctggc actctgggtt gctagccccg ccgggactg  
 121 ggccctcagac actgcgcggt tccctcggag cagcaagcta aagaaagccc ccagtgcggg  
 181 cgaggaaagg ggcggcgggg aaagatgcgc ggcgttggct ggcagatgct gtcctctgtc  
 5 241 ctgggggttag tgctggcgat cctgaaccaag gtggcaccgc aggcgtgccc ggcgcagtgc  
 301 tcttgctcgg gcagcacagt ggactgtcac gggctggcgc tgcgcagcgt gccaggaat  
 361 atcccccgca acaccgagag actggattta aatggaaata acatcacaag aattacgaag  
 421 acagattttg ctgggtcttag acatctaaga gttcttcagc ttatggagaa taagattagc  
 481 accattgaaa gaggagcatt ccaggatctt aaagaactag agagactgcg tttaaacaga  
 10 541 aatcaccttc agctgtttcc tgagttgctg tttcttggga ctgcgaagct atacaggctt  
 601 gatctcagtg aaaaccaaatt tcaggcaatc ccaaggaaag ctttccgtgg ggcagttgac  
 661 ataaaaaatt tgcaactgga ttacaaccag atcagctgta ttgaagatgg ggcattcagg  
 721 gctctccggg acctggaagt gctcactctc aacaataaca acattactag actttctgtg  
 781 gcaagtttca accatatgcc taaacttagg acttttcgac tgcattcaaa caacctgtat  
 15 841 tgtgactgcc acctggcctg gctctccgac tggcttcgcc aaaggcctcg ggttggctctg  
 901 tacactcagt gtatgggccc ctcccacctg agaggccata atgtagccga ggttcaaaaa  
 961 cgagaatttg tctgcagtggt tcaccagtca tttatggctc cttctttagt tgttttgcac  
 1021 tgccctgccc cctgtacctg tagcaacaat atcgtagact gtcgtgggaa aggtctcact  
 1081 gagatcccca caaatcttcc agagaccatc acagaaatac gtttgaaca gaacacaatc  
 20 1141 aaagtcattc ctcttgagc tttctcacca tataaaaagc ttagacgaat tgacctgagc  
 1201 aataatcaga tctctgaact tgcaccagat gctttccaag gactacgctc tctgaattca  
 1261 cttgtctctc atggaaataa aatcacagaa ctccccaaaa gtttatttga aggactgttt  
 1321 tccttacagc tcctattatt gaatgccaac aagataaact gccttcgggt agatgctttt  
 1381 caggatctcc acaacttgaa ccttctctcc ctatatgaca acaagcttca gaccatcgcc  
 25 1441 aaggggacct tttcacctct tcggggccatt caaactatgc atttggccca gaaccccttt  
 1501 atttctgact gccatctcaa gtggctagcg gattatctcc ataccaaccc gattgagacc  
 1561 agtgggtgcc gttgcaccag ccccgccgc ctggcaaaaca aaagaattgg acagatcaaa  
 1621 agcaagaaat tccgttgttc agctaaagaa cagtatttca tccaggtac agaagattat  
 1681 cgatcaaaat taagtggaga ctgctttgcg gatctggctt gccctgaaaa gtgtcgctgt  
 30 1741 gaaggaaaca cagtagattg ctctaataca aagctcaaca aaatcccga gcacattccc  
 1801 cagtacactg cagagttgcg tctcaataat aatgaattta ccgtgttggg agccacagga  
 1861 atctttaaga aacttctca attacgtaaa ataaacttta gcaacaataa gatcacagat  
 1921 attgaggagg gagcatttga aggagcatct ggtgtaaatg aaatacttct tacgagtaat  
 1981 cgtttgaaaa atgtgcagca taagatgttc aagggttgg aaagcctcaa aactttgatg  
 35 2041 ttgagaagca atcgaataac ctgtgtgggg aatgacagtt tcataggact cagttctgtg  
 2101 cgtttgcttt ctttgtatga taatcaaatt actacagttg caccaggggc atttgatact  
 2161 ctccattctt tatctactct aaacctcttg gccaatcctt ttaactgtaa ctgctacctg  
 2221 gcttgggttg gagagtggct gagaaagaag agaattgtca cgggaaatcc tagatgtcaa  
 2281 aaaccatact tctgaaaga aatacccatc caggatgtgg ccattcagga cttcacttgt  
 40 2341 gatctagcaa ttggagccaa tagttgtctc ccactttctc gctgtcctac tgaagtact  
 2401 tgcttgata cagtcgtccg atgtagcaac aagggttga aggtcttgcc gaaaggtatt  
 2461 ccaagagatg tcacagagtt gtatctggat ggaaaccaat ttacactggg tccaaggaa  
 2521 ctctccaact acaaacattt aacacttata gacttaagta acaacagaat aagcacgctt  
 2581 tctaatacaga gcttcagcaa catgaccag ctctcacct taattcttag ttacaaccgt  
 45 2641 ctgagatgta ttcctcctcg cacctttgat ggattaaagt ctcttegatt actttctcta  
 2701 catggaaatg acatttctgt tgtgcctgaa ggtgctttca atgatcttc tgcattatca  
 2761 catctagcaa ttggagccaa cctcttttac tgtgattgta acatgcagtg gttatccgac  
 2821 tgggtgaagt cggaatataa ggagcctgga attgctcgtt gtgctggctc tggagaaatg  
 2881 gcagataaac ttttactcac aactccctcc aaaaaattta cctgtcaagg tctgtggat  
 50 2941 gtcaatatcc tagctaagtg taaccctgct ctatcaaatc cgtgtaaaaa tgatggcaca  
 3001 tgtaatatgt atccagttga cttttaccga tgcacctgtc catatggttt caaggggcag  
 3061 gactgtgatg tcccaattca tgcctgcac agtaacccat gtaaacatgg aggaacttgc  
 3121 ca:ttaaagg aaggagaaga agatggattc tgggtgattt gtgctgatgg atttgaagga  
 3181 gaaaattgtg aagtcaacgt tgatgtattg gaagataatg actgtgaaaa taattctaca  
 55 3241 tgtgtcgatg gcattaataa ctacacatgc ctttgccac ctgagtatac aggtgagttg  
 3301 tgtgaggaga agctggactt ctgtgcccag gacctgaacc cctgccagca cgattcaag  
 3361 tgcatcctaa ctccaaaggg attcaaatgt gactgcacac cagggtacgt aggtgaacac  
 3421 tgcgacatcg attttgacga ctgccaaagac aacaagtgtg aaaacggagc cactgcaca  
 3481 gatgcagtga acggctatac gtgcatatgc cccgaagggt acagtggctt gttctgtgag  
 60 3541 ttttctccac ccatggctct ccctcgtagc agccctgtg ataattttga ttgtcagaat  
 3601 ggagctcagt gtatcgtcag aataaatgag ccaatatgtc agtgtttgcc tggctatcag

3661 ggagaaaagt gtgaaaaatt ggtagtggtg aattttataa acaaagagtc ttatcttcag  
 3721 attccttcag ccaaggttcg gcctcagacg aacataacac ttcagattgc cacagatgaa  
 3781 gacagcggaa tcctcctgta taaggggtgac aaagaccata tcgcggtaga actctatcgg  
 3841 gggcgtgttc gtgccagcta tgacaccggc tctcatccag cttctgccat ttacagtgtg  
 5 3901 gagacaatca atgatggaaa cttccacatt gtggaactac ttgccttggg tcagagtctc  
 3961 tctttgtccg tggatgggtg gaaccccaaa atcatcacta acttgtcaaa gcagtccact  
 4021 ctgaattttg actctccact ctatgtagga ggcattgccg ggaagagtaa cgtggcatct  
 4081 ctgcgccagg cccctgggca gaacggaacc agcttcacg gctgcatccg gaacctttac  
 4141 atcaacagtg agctgcagga cttccagaag gtgccgatgc aaacaggcat tttgcctggc  
 10 4201 tgtgagccat gccacaagaa ggtgtgtgcc catggcacat gccagcccag cagccaggca  
 4261 ggcttcacct gcgagtcca ggaaggatgg atggggcccc tctgtgacca acggaccaat  
 4321 gacccttgcc ttgaaataa atgcgtacat ggcacctgct tgcccatcaa tgcgttctcc  
 4381 tacagctgta agtgcttggg agggccatgga ggtgtcctct gtgatgaaga ggaggtctg  
 4441 ttttaacccat gccaggcgat caagtgcagg cacgggaagt gcaggccttc aggtctgggg  
 15 4501 cagccctact gtgaatgcag cagtggatac acgggggaca gctgtgatcg agaaatctct  
 4561 tgtcgagggg aaaggataag agattattac caaaagcagc agggctatgc tgcttgccaa  
 4621 acaaccaaga aggtgtcccg attagagtgc agaggtgggt gtgcaggagg gcagtgtgt  
 4681 ggaccgctga ggagcaagcg gcggaatac tctttcgaat gcactgacgg ctctccttt  
 4741 ttggcagagg ttgagaaagt ggtgaagtc ggctgtacga ggtgtgtgtc ctaaacacac  
 20 4801 tccggcagc tctgtctttg gaaaagggtg tatacttctt gaccatgtgg gactaatgaa  
 4861 tgcttcatag tggaaatatt tgaatatat tgtaaaatac agaacagact tatttttatt  
 4921 atgagaataa agactttttt tctgcatttg

SEQ ID NO: 229

25 Amino acid sequence of human SLIT2 encoded by the DNA sequence shown in SEQ ID NO:  
 228.

MRGVGWQMLSLSLGLVLAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL  
 DLNGNNITRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRLNRNHLQLPPE  
 LLFLGTAKLYRLDLSENQIQAI PRKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVL  
 30 TLNNNNITRLSVASFNMPLKRTFRLHSNNLYCDCHLAWLSDWLRQRPRVGLYTQCMGPS  
 HLRGHNAVEVQKREFVCSGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPE  
 TITEIRLEQNTIKVIPPGAFSPYKLRRLDLSNNQISELAPDAFQGLRSLNSLVLYGNKI  
 TELPKSLFEGFLFSLQLLLLNANKINCLRVDAFQDLHNLNLSLYDNKLQTIAGKTFSPLR  
 AIQTMHLAQNPFCIDCHLKLWADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSA  
 KEQYFIPGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPEHIPQYTAE LRL  
 35 NNNEFTVLEATGIFKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLENVQHK  
 MFKGLESILKTLMLRSNRITCVGNDSFIGLSSVRLLSLYDNQITTVAPGAFDTLHLSLSTLN  
 LLANPFNCNCTYLAWLGEWLKRIKRVGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDNS  
 CSPLSRCPTECTCLDTPVRC SNKGLKVLPGKIPRDVTELYLDGNQFTLVPKELS NYKHLT  
 40 LIDLNNRISTLSNQSFNMTQLLTLLSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVV  
 PEGAFNDLSALSHLAIGANPLYCDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLTT  
 PSKKFTCQGPVDVNI LAKCNPCLSNPCKNDGT CNSDPVDFYRCTCPYGFKGQDCDVPIHA  
 CISNPCKHGGTCHLKEGEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNY  
 TCLCPPEYTGELCEEKLDFAQDLNPCQHDSKILTPKGFKCDCTPGYVGEHCDIDFDDC  
 QDNKCKNGAHTDAVNGYTCICPEGYSGLFCBFSPPMVLPRTSPCDNFDCQNGAQCI VRI  
 45 NEPICQLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYK  
 GDKDHI AVELYRGRVRASYDTGSHPASAIYSVETINDGNFHVIVELLALDQSLSLSV DGGN  
 PKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCIRNLYINSELQDF  
 QKVPMTQILPGCEPCHKVKCAHGTCPSSQAGFTCECQEGWMGPLCDQRTNDPCLGNKC  
 VHGTCLPINA FSYSCLEBGHGGVLCDEEDLFNPCQAIKCKHKGKRLSGLGQPYCECSS  
 50 GYTGDSCDREISCRGERIRDYYQKQGYAACQTTKVSRL ECRGGCAGGQCCGPLRSKRR  
 KYSFECTDGSFVDEVEKVVKCGCTRCVS

SEQ ID NO: 230

Amino acid sequence of human SLIT2, a soluble active secreted form derived from SEQ ID  
 NO:229.



QCSCSGSTVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHLRVLQLMENK  
 ISTIERGAFQDLKELERLRNLRNHLQFLPELLPLGTAKLYRLDLSENQIQAI PRKAFRGA  
 VDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNMHPKLRTFRLHSNN  
 LYCDCHLAWLSDWLRQRPVGLYTQCMGPSHLRGHNVAEVQKREFVCSGHQSFMAPSCSV  
 5 LHCPAACTCSNNIVDCRGKGLTEIPTNLPETITIRLEQNTIKVIPPGAFSPYKKLRRID  
 LSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGFLPSLQLLLLNANKINCLRVD  
 AFQDLHNLNLLSLYDNKLQTIAGKTFSPLRAIQTMHLAQNPFCDCHLKWLADYLHTNPI  
 ETSGARCTSPRRLANKRIGQIKSKKFRCSAKEQYFIPGTEDYRSKLSGDCFADLACPEKC  
 RCEGTTVDCSNQKLNKIPHIPIQYTAELRLNNEFTVLEATGIFKKLPQLRKINFSNNKI  
 10 TDIREGAFEGASGVNEILLTSNRLENVQHKMFKGLESKTLMLRSNRITCVGNDSSFGLS  
 SVRLLSLYDNQITTVAPGAFDTLHSLSTLNLNLANPFCNCYLAWLGEWLKKRIVTGNPR  
 CQKPYFLKEIPIQDVAIQDFTCDGNDNDSCLSRCPTECTCLDTVVRCNKGKLVLPK  
 GIPROVTELYLDGNQFTLVPKELSNIYKHLTLIDLSNNRISTLSNQSFNMTQLLTLILSY  
 NRLRCIPPRTFDGLKSLRLLSLHGNDISVVEGAFNDLSALSHLAIGANPLYCDCNMQWL  
 15 SDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKCNPCLSNPCKND  
 GTCNSDPVDFYRCTCPYGFKGQDCDVPFHACISNPCKHGGTCHLKEGBEDGFWCICADGF  
 EGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDCAQDLNPCQHD  
 SKCILTPKGFKDCDCTPGYVGEHCDIDFDDCQDNKCKNGAHCTDAVNGYTCICPEGYSGLF  
 CEFSPPMVLPR

20 SEQ ID NO: 231

gi|30794373|ref|NM\_178804.2| Mus musculus slit homolog 2 (Drosophila) (Slit2), mRNA

1 tattcagaac ttaagttgcc cacggatcct ctgctctgct agaaaggctt gaagagcaga  
 61 ggaaagacct gtgcttctgt cagctctcct gccccatata actgttccag attactgtgt  
 121 gagcatctcc ccgggtgctg tgggctgcaa gccacagccc aggcactggg cctcggacac  
 25 181 tgcccgggtt ttacacaacc gaaagctcaa gagaagtcct tcaaagcaag gagtctatta  
 241 gggaagatga gtggcattgg ctggcagaca ctgtccctat cgtcgggggtt agtgtgtcgc  
 301 atcttgaaca aggtggcgcc gcaggcgtgc ccggcccagt gctcctgttc aggcagcacg  
 361 gtggactgtc atgggctggc actgcgagcgt gtgcccagga atatcccccg caacaccgag  
 421 agactggatt tgaatggaaa taacatcacg aggatcacga agacagattt tgctgggtctc  
 30 481 aggcacctca gagttcttca gctcatggag aacagaatca gcaccatcga gaggggagca  
 541 ttccaggatc ttaaggagct ggaaagactg cgtttaaaca gaaataacct tcagtgtgtt  
 601 cctgagctgc tgtttctcgg gactgcgaag ctctaccggc ttgatctcag tgaaaaatcaa  
 661 attcaagcaa ttccaaggaa ggctttccgt ggggcagttg acattaaaaa cctgcaactg  
 721 gattacaacc agatcagctg cattgaagat ggggcgttca gagctctacg agatctggaa  
 35 781 gtgctcactc tgaacaataa caatattact agactttcag tggcaagttt caaccatattg  
 841 cctaaactta ggacatttcg actccactcg aacaacttgt actgcgactg ccacctagcc  
 901 tggctctcag actggcttcg ccaaaggcca cgggtgggct tgtacactca gtgtatgggc  
 961 ccatcccacc tgagggggcca caatgtagca gaggttcaaa aacgagagtt tgtctgagc  
 1021 ggtcaccagt cattcatggc tccctcctgc agtgtgctgc actgccccgc tgcctgtacc  
 40 1081 tgtagcaaca acattgtaga ctgccgaggg aaaggtctca ctgagatccc cacaatctg  
 1141 cctgagacca tcacagaaat acgtttggaa cagaactcca tcagggtcat cctccagga  
 1201 gccttctcac catacaaaaa gcttagacga ctagacctga gcaacaacca gatctctgaa  
 1261 cttgcaccag atgccttcca aggactgcgc tctctgaatt cacttgtcct gtatggaaat  
 45 1321 aaaatcacag aactcccaaa aagtttattc gaaggactat ttctcttgca gctactatta  
 1381 ttgaatgcc acaagataaa ctgccttcgg gtagatgctt ttcaggacct gcacaacttg  
 1441 aaccttctct cttatatga caataagctt cagacgggtg ccaagggcac cttctcagcc  
 1501 ctgagagcca tccaaactat gcatttgccc cagaatcctt tcatttgtga ctgccatctc  
 1561 aagtggctag cggattatct ccacaccaac ccaattgaga ccagcgggtg ccgttgacc  
 1621 agcccccgcc gcctggcaaa caaaagaatt ggacagatca aaagcaagaa attccggtgt  
 50 1681 tcaggtagac aagattatcg atcaaaatta agtggagact gctttgcaga cttggctgtg  
 1741 cctgagaagt gtcgctgtga agggaccaca gtagactgct ccaatcaaa actcaaaaa  
 1801 atccctgacc atattcccca gtacacagca gagctgcgtc tcaataataa tgaattcaca  
 1861 gtgttagaag ccacgggaat atttaagaaa ctctctcagt tacgtaaaat caacttttagc  
 1921 aacaataaga tcacggatat cgaggagggt gcatttgaag gcgcgtctgg tgtgaatgaa  
 55 1981 attcttctca ccagtaaccg tttggaaaat gttcagcata agatgttcaa aggactggag  
 2041 agcctcaaaa cattgatgct gagaagtaat cgaataagct gtgttgggaa cgacagtttc  
 2101 ataggactcg gctctgtgcg tctgctctct ttatatgaca atcaaatatc cacagtggca

2161 ccaggagcat ttgattctct ccattcatta tccactctaa acctcttggc caatcctttc  
 2221 aactgttaact gtcacctggc atggctggga gaatggctca gaagggaaa aattgttaaca  
 2281 ggaaatcctc gatgccaaaa accctacttc ctgaaggaaa tcccaatcca ggatgtagcc  
 2341 attcaggact tcacctgtga tgatggaaat gatgacaata gttgctctcc actctcccgt  
 5 2401 tgtccttctg aatgtacctg cttggataca gtggtacgat gtagcaacaa gggcttgaag  
 2461 gttttgccta aaggatttcc aaaagatgtc acagagctgt atctggatgg gaaccagttt  
 2521 acgctgggtcc cgaaggaaact ctctaactac aaacatttaa cacttataga cttaagtaac  
 2581 aaccgaataa gcaccctttc caatcaaacg ttcagcaaca tgaccagct tctcacctta  
 2641 atcctcagtt acaaccgtct gagatgtatc cctccacgaa cctttgatgg attgaagtct  
 10 2701 cttcggttac tgtctttaca tggaaatgac atttctgttg tgcctgaagg tgccttcaat  
 2761 gacttgtcag ccttgctaca cttagcgatt ggagccaacc ctctttactg tgattgtaac  
 2821 atgcagtggt tatccgactg ggtgaagtgc gaatataagg aacctggaat tgcacgtgt  
 2881 gccggccctg gagaaatggc agataaatta ttactcacta ctccctccaa aaaatttaca  
 2941 tgtcaaggtc ccgtggatat cactattcaa gccaaagtga atccctgctt atcaaatcca  
 15 3001 tgtaaaaatg atggcacctg taacaatgac cccgttgatt tttatcgatg tacctgcccc  
 3061 tatggattca agggtcagga ctgtgatgtc cccattcatg cttgtatcag taatccatgt  
 3121 aaacatggag gaacttgta cttaaaggaa ggagagaatg ctggattctg gtgcacttgt  
 3181 gctgatgggt ttgaaggaga aaactgtgaa gtcaatattg atgattgtga agataatgat  
 3241 tgtgaaaata attctacatg cgttgatgat attaacaact acacatgtct ttgccaccg  
 20 3301 gaatacacag gtgaaatggt taggaaaaag ctggacttct gtgcacaaga cttgaatccc  
 3361 tgccagcatg actccaagtg catcctgact ccaaagggat tcaagtgtga ctgcactcca  
 3421 ggatacattg gtgagcactg tgacattgac tttgatgact gccaagataa caagtgtaaa  
 3481 aacggtgctc actgacacaga tgccgtgaac ggatacacgt gcgtctgtcc tgaaggctac  
 3541 agtggcttgt tctgtgagtt ttctccaccc atggctcctc ctgcaccag cccctgtgat  
 25 3601 aattttgatt gccagaatgg agcccagtgt atcatcagga taaatgaacc aatatgccag  
 3661 tgtttgctg gctacctggg agagaagtgt gagaaattgg tcagtgtgaa ttttgtaaac  
 3721 aaagatgcct atcttcagat tccttcagcc aagggtcggc ctccagacaaa catcacactt  
 3781 cagattgcca cagatgaaga cagcggcatc ctctgtata aagggtgacaa agaccacatt  
 3841 gccgtggaac tctatagagg gcgagttcga gccagctatg acaccggctc tcatccggct  
 30 3901 tctgccattt acagtgtgga gacaatcaat gatggaaact tccacattgt ggagctactg  
 3961 accctggatt ccagtctttc cctctctgtg gatggaggaa gccctaaagt catbccaat  
 4021 ttgtcaaaa aatctactct gaatttcgac tctccactct atgtaggagg catgcccggg  
 4081 aaaaaataacg tggcatccct gcgccaggcc cctgggcaaa atggcaccag tctccatggc  
 4141 tgtatccgga acctttacat tcacagttag ctgcaggact tccggaaaaat gcctatgcaa  
 35 4201 accggaattc tgctggctg tgaaccatgc cacaagaaag tatgtgcca tggcatgtgc  
 4261 cagcccagca gccaatcagg cttcacctgt gaatgtgagg aagggtggat ggggcccctc  
 4321 tgtgaccaga gaaccaatga tccttgctc ggaaacaaat gtgtgcatgg gacctgctg  
 4381 cccatcaatg ccttctccta tagttgcaag tgctggagg gccatggcgg tgcctctgt  
 4441 gatgaagaag aagatctctt taaccctgtc cagatgatca agtgcaagca tgggaagtgc  
 40 4501 aggctttctg gagtgggcca gccctattgt gaatgcaaca gtggattcac cggggacagc  
 4561 tgtgatagag aaatttcttg tcgaggggaa cggataaagg actattacca gaagcagcag  
 4621 gggtacgctg cctgtcaaac aactaagaaa gtatctcgct tggaaatgcag aggcgggtgc  
 4681 gctggaggcc agtgcgtgtg acccttgaga agcaagaggc ggaaatactc tttcgaatgc  
 4741 acagatggct cctcatttgt ggacgaggtt gagaaagtgg tgaagtgcgg ctgcgcgaga  
 45 4801 tgtgcctcct aagcgcgtct ctagaagctt ctagcttcgg cgaaggttgt acacttcttg  
 4861 accatggttg actaattcat gcttcataat ggaaatattt gaaatatatt gtaaaatata  
 4921 gaacagactt atttttatta tgataataaa gacttgctg catttgga aaataataa  
 4981 taaaagccac gcttgacta aagctcccc tacactggag aagtgtggag aaaagccaca  
 5041 cttggaggca ttagtgaagc ggtgggtacc attgcaacac ggagccatct ttggaacacg  
 50 5101 ccaatactag cagaagcaca tctacaagag cctgacatgg gactgtacgc tcatatgcgc  
 5161 ttgcctagag cgtcaaaccg gtgacccctc tcacatcagt tccaaggac acatcgagca  
 5221 caagtgtgtg agtgagactg agaggtagat gaatgggaat ctagaatctg tagataggtc  
 5281 caagggatga gatataatgt gtgcaagatc taaagtgtcc tgacatctgg gtttcattcc  
 5341 gaaagaaatg ggtgtggagt cgtgtgatgt attttattgt ctttggaat gacagcgcac  
 55 5401 aggcagaaac agcacacgaa ggtgttttagc taccagtttc cagtattaaa tttttgttaa  
 5461 tataaatgac aaaggagatg ataaagaacc aatagattat tgataaataa aaattagtaa  
 5521 taatatgaat ttttgtttct atgagttcta aatatcccta tacagtattc attttcaatg  
 5581 aggaatatatt attgtattaa agtacatttt acttaacgat ttagatcatc attttggact  
 5641 gtttcttgat gctttaatat atattaattt ataattatcc tgatattttt gtacattttc  
 60 5701 aaactttaaa aaaaaatcag gatttttttg gcaatagtac taacatagga tgttattttg  
 5761 ttatctcggg ataagtatat gttgttctgt tggtgcacct tgatgtctga ccattgatgc

5821 cactgaactg tggcctcctt caggacactt gcagaagggt tgcaaagtct cagagtagaa  
 5881 acagcaagtg aatctatctg ccacatgtcc tcacaagaag aaagaactta tctgtagcaa  
 5941 tggccagtaa gaagtagatg ttttacaat taattaataa acatttcctt gccatgcttt  
 6001 gttggggagg gtgggagaag aaagactgtt atgtcaaaaa ttgtaatttc tttccaaaca  
 5 6061 ataaaattcc tttattaacg taa

SEQ ID NO: 232

Amino acid sequence of mouse SLIT2 encoded by the DNA sequence shown in SEQ ID NO: 231.

10 MSGIGWQTLSSLGLVLSILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL  
 DLNGNNITRITKTDFAGLRHLRLVQLMENRISTIERGAFQDLKELERLRNRRNNLQLFPE  
 LLFLGTAKLYRLDSENQIQAIIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVL  
 TLNNNNITRLSVASFNHMPKLRTRFRLHSNNLYCDCHLAWLSDWLRQRPRVGLYTQCMGPS  
 HLRGHNVAEVQKREFVCSGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPE  
 15 TITEIRLEQNSIRVIPPGAFSPYKKLRRLLDLSNNQISELAPDAFQGLRSLNSLVLYGNKI  
 TELPKSLFEGFLSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTVAKGTFSSALR  
 AIQTMHLAQNPFCIDCHLKWLDLYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSG  
 TEDYRSKLSGDCFPADLACPEKCRCEGTTVDCSNQRNLNIPDHIPQYTAELRLNNNEFTVL  
 EATGIFKKLPQLRKINFSSNNKITDIEGAFEGASGVNEILLTSNRLENVQHKMFKGLES  
 20 KTLMLRSNRISCVGNDSPFGLGSVRLLSLYDNQITTVAPGAFDSLHSLSTLNLLANPFNC  
 NCHLAWLGEWLRRKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDNNSCSPLSRCP  
 SECTCLDTVVRCSNKGKLVLPKGI PKDVTLEYLDGNQFTLVPKELSNYKHLTLIDLSNNR  
 ISTLSNQSFNSMTQLLTLLSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVEGAFNDL  
 SALSHLAIGANPLYCDNCMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQ  
 25 GPVDITIQAKCNPCLSNPKNDGTGNDPVDVDFYRCTCPYGFKGQDCDVP IHACISNPCKH  
 GGTCHLKEGENAGFWCTCADGFEGENCEVNIDDCEDNDCENNSTCVDGINNYTCLCPPEY  
 TGELCEEKLDFAQDLNPCQHDSKCILTPKGFKCDCTPGYIGEHCDIDFDDCQDNKCKNG  
 AHCTDAVNGYTCVCEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIIRINEPICQCL  
 PGYLGEKCEKLVSVNFVNKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAV  
 30 ELYRGRVRSYDGTGSHPASAIYSVETINDGNFHVIELLTLDSLSLSDVGGSPKVIITNLS  
 KQSTLNFDSPLYVGGMPGKNNVASLRQAPGQNGTSFHGCIRNLYINSELQDFRKMMPQTG  
 ILPGCEPCHKKVCAGHMCQPSQSGFTCECEEGWMPGLCDQRTNDPCLGNKCVHGTCLPI  
 NAFSYSCKCLEGGHGVLCDEEDLFPNCPQMIKCKHGKRLSGVGGPYCECNSGFTGDS  
 REISCRGERIRDYYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTD  
 GSSFVDEVEKVVKCGCARCAS

35 SEQ ID NO: 233

ENSRNOT00000005477 cDNA sequence, EnsEMBL transcript [Rattus norvegicus]

1 atgagtggca ttggctggca gacactgtcc ctatctcttg cgtagtggt gtcgactctg  
 61 aaccaggtgg cgcctcaggc gtgcccggcc cagtgtcctt gttcaggcag cacagtggac  
 121 tgcctatgggc tggcactgcg cagtgtgccc aggaatatcc cccgcaacac ggagagactg  
 40 181 gattttgaatg gaaataacat cacaaggatc acgaagacag attttgcggg tctcagacac  
 241 ctcagagttc ttcagctcat ggagaacaag atcagcacca tcgagagggg agcattccag  
 301 gatctttaagg agctagaaag actgcgttta aacagaaata accttcagtt gtttctctgag  
 361 ctgctgtttc ttgggactgc gaagctctac cggcttgatc tcagtgaataa tcagattcaa  
 421 gcaattccaa ggaaggcttt ccgtggtgca gttgacatta aaaatctgca gttggattac  
 45 481 aaccagatca gctgcattga agatggggca ttccgagctc tgcgagatct ggaagtgtct  
 541 actctgaaca ataacaatat tactagactt tcagtggcaa gtttcaacca tatgcctaaa  
 601 cttaggacat ttcgactcca ctccaacaac ctatactgcg actgccacct ggcctggctc  
 661 tgggactggc ttcgcccagg gccacgggtg ggcttgata ctcagtgtat gggcccatcc  
 721 cacctgaggg gccataatgt agcagaggtt caaaaacgag agtttgtctg cagtggtcac  
 50 781 cagtcattca tggctccctc ctgcaggtg ctgcactgcc cgattgcttg tacctgtagc  
 841 aacaacattg tagactgccg agggaaaggt ctcactgaga tccccacaaa tctgctgag  
 901 accatcacag aaatacgttt ggaacagaac tccataaggg tcatccctcc aggagcattc  
 961 tcaccatata aaaagcttcg acgactagac ctgagtaata accagatctc ggaacttgct

1021 ccagatgcct tccaaggact gcgttctctg aattcccttg tcctgtatgg aaataaaatc  
 1081 acagaactcc caaaaagttt atttgaagga ctgttttccct tacagctact attattgaat  
 1141 gccacaaga taaactgcct tcgggtagat gcttttcagg acctgcacaa cttgaacctt  
 1201 ctctccttat acgacaataa gcttcagact gttgccagg gcaccttctc agctctcaga  
 5 1261 gccatccaaa ctatgcattt ggcccagaat cctttcattt gtgactgcca tctcaagtgg  
 1321 cttagcggatt atctccacac caacccaatt gagaccagcg gtgcccgttg caccagtcctc  
 1381 cgccgcctgg ctaacaaaag aattggacag atcaaaagca agaaattccg ttgttcagggt  
 1441 acagaagatt atcgatcaaa attaagtggg gactgctttg cagacttggc ttgtcctgaa  
 1501 aaatgtcgct gtgaagggac cacagtagac tgctccaatc aaaaactcaa caaaatccca  
 10 1561 gaccatattc cccagtacac agcagagctg cgtctcaata ataataaatt cacagtgtta  
 1621 gaagccacgg gaatttttaa gaaacttcct caattgcgta aaatcaacct tagcaacaat  
 1681 aagatcacctg atatcgagga gggggcattc gaagggtgctg ctgggtgtgaa tgattgtctg  
 1741 cttaccagta accgttttga aaatgttcag cataagatgt tcaaaggatt ggagagcctc  
 1801 aaaacattga tgctgagaag taatcgaata agctgtgtgg gaaacgacag ttccacagga  
 15 1861 ctcggttctg tgcgtctgct ctctttatat gacaatcaaa ttaccacagt tgcaccagga  
 1921 gcatttggta ctctccattc attatctaca ctaaacctct tggccaatcc tttcaactgt  
 1981 aactgtcacc tggcatggct tggagaatgg ctccagaagg aaagaattgt aacaggaaat  
 2041 cctcgatgcc aaaaacccta ctctttgaag gaaataccaa tccaggatgt agccattcag  
 2101 gacttcacctg gtgatgacgg aaacgtatgt aatagctgct ctccactctc ccgttgtcct  
 20 2161 tcggaatgta cttgcttggg tacagttagta cgatgtagca acaagggtct gaagggtctta  
 2221 cctaaaggca ttccaagaga tgtcacagaa ctgtatctgg atgggaacca gtttacactg  
 2281 gtcccgaagg aactctccaa ctacaaacat ttaacactta tagacttaag taacaacaga  
 2341 ataagcacc tttccaacca aagcttcacg aacatgaccc aacttctcac cttaattctc  
 2401 agttacaacc gtctgagatg tatccctcca cggacctttg atggattgaa atctcttctg  
 25 2461 ttactgtctc tacatggaaa tgacatttct gtcgtgcctg aagggtgcctt tgggtgacct  
 2521 tcagccttgt cacacttagc aattggagcc aacctctttt actgtgattg taacatgcag  
 2581 tggttatccg actgggtgaa gtcggaatat aaggaacctg gaattgcccg ctgtgcccgt  
 2641 cccggagaaa tggcagataa attgttactc acaactccct ccaaaaaatt tacatgtcaa  
 2701 ggtcctgtgg atgttactat tcaagccaag tgtaaccctt gcttgtcaaa tccatgtaaa  
 30 2761 aatgatggca cctgtaacaa tgacccgggtg gatttttatc gatgcacctg cccatagtgt  
 2821 ttcaagggcc aggactgtga tgtccctatt catgcctgta tcagtaatcc atgtaaacat  
 2881 ggaggaactt gccacttaa agaaggagag aatgatggat tctggtgtac ttgtgtgtat  
 2941 gggtttgaag gagaaaactg tgacatcaat attgatgatt gcgaagataa tgatttgtaa  
 3001 aataattcta catgcgttga tggaaattaac aactacacgt gtctttgccc accggaatac  
 35 3061 acaggcgaac tgtgtgagga aaaactggac ttctgtgcac aagacctgaa tccctgccag  
 3121 catgactcca agtgatcctt gacgccaaag ggattcaagt gtgactgcac tccgggatac  
 3181 attggtgagc actgtgacat cgactttgat gactgccaag ataacaagtg caaaaacggt  
 3241 gctcattgca agatgcagt gaacggatac acatgtgtct gtcctgaagg ctacagtggc  
 3301 ttgttctgtg agttttctcc acctatgggt ctcctcgca ccagcccctg tgataatttt  
 40 3361 gattgtcaga atggagccga gtgtatcatc aggggtgaatg aaccaatatg ccaggtgttg  
 3421 cctggctact tgggagagaa gtgtgagaaa ttgggtcagt tgaattttgt aaacaaagag  
 3481 tccatctctc agattccttc agccaagggt cgacctcaga caaacatcac acttcagatt  
 3541 gccacagatg aagacagcgg catcctcttg tacaagggtg acaaggacca cattgctgtg  
 3601 gaactctatc gagggcgagt tcgagccagc tatgacaccg gctctcaccg ggcttctgcc  
 45 3661 atttacagtg tggagacaa caatgatgga aacttcacaa ttgtagagct actgacctg  
 3721 gattcgagtc ttccctctc tgtggatgga ggaagcccta aaatcatcac caatttgtca  
 3781 aaacaatcta ctctgaattt cgactctcca cttacgtag gaggtatgcc tgggaaaaat  
 3841 aacgtggctt cgctgcgcca ggcccctggg cagaacggca ccagcttcca tggctgtatc  
 3901 cggaaccttt acattaacag tgaactgcag gacttccgga aagtgcctat gcaaacggga  
 50 3961 attctgcctg gctgtgaacc atgccacaag aaagtgtgtg cccatggcac atgccagccc  
 4021 agcagccaat caggcttcac ctgtgaatgt gaggaagggt ggatggggcc cctctgtgac  
 4081 cagagaacca atgatccctg tctcggaaac aaatgtgtac atgggacctg cttgccatc  
 4141 aacgccttct cctacagctg caagtgcctg gagggccacg gcggggctct ctgtgatgaa  
 4201 gaagaagatc tgtttaaccc ctgccaggtg atcaagtgca agcacgggaa gtgcaggctc  
 55 4261 tctgggctcg ggcagcccta ttgtgaatgc agcagtggat tcaccgggga cagctgtgac  
 4321 agagaaattt cttgtcgagg ggaacggata agggattatt accaaaagca gcagggttac  
 4381 gctgcctgtc aaacgactaa gaaagtatct cgcttggagt gcagaggcgg gtgtgctggg  
 4441 gggcagtgct gtggacctct gagaagcaag aggcgggaaat actcttctga atgcacagat  
 4501 ggatcttcat ttgtggacga ggtcgagaag gtggtgaagt gcggctgcac gagatgtgcc  
 60 4561 tcctaa

## SEQ ID NO: 234

Amino acid sequence of rat SLIT2 encoded by the DNA sequence shown in SEQ ID NO: 233.

```

MSGIGWQTLSSLALVLSILNQVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRINTERL
DLNGNNITRITKTDFAGLRHLRVLQLMENKISTIERGAFQDLKELERLRNLRNQLFPE
5  LLFLGTAKLYRLDLSNQIQAIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVL
TLNNNNITRLSVASFNMHPKLRTFRLHSNNLYCDCHLAWLSDWLRQRPRVGLYTQCMGPS
HLRGHNVAEVQKREFVCSGHQSFMAPSCSVLHCPACTCSNNIVDCRGKGLTEIPTNLPE
TITEIRLEQNSIRVIPPGAPSPYKLRRLDLSNNQISELAPDAFQGLRSLNSLVLYGNKI
TELPKSLFEGFLFSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTVAKGTFSALR
10 AIQTMHLAQNPFICDCHLKWLDLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSG
TEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPDHIPOYTAEHLNNNEFTVL
EATGIFPKLPQLRKINLSNNKITDIEGAFEGASGVNEILLTSNRLNVQHKMFKGLES
KTLMLRSNRISCVGNDSTGLGSRVLLSLYDNQITTVAPGAFGTLHSLSTLNLNPFNC
NCHLAWLGEWLRKRIVTGNPRCQKPYFLKBIPIQDVAIQDFTCDDGNDNDCSPLSRCP
15 SECTCLDTVVRCNSIRVIGLVPKGI PRDVTLEYLDGNQFTLVPKELSNYKHLTLIDLSNNR
ISTLSNQSFNSMTQLTLILSYNRLRCIPRRTFDGLKSLRLLSLHGNDISVVPEGAFGDL
SALSHLAIGANPLYCDNMQWLSDWKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQ
GPVDVTIQAKCNPCLSNPKNDGTCNNDPVDIFYRCTCPYGFKGQDCDVPFHACISNPKH
GGTCHLKEGENDGFWCTCADGFEGESCDINIDDCEDNDCENNSTCVDGINNYTCLCPPEY
20 TGELCEEKLDFAQDLNFCQHDSCILTPKGFKCDCTPGYIGEHCDIDFDDCQDNKCKNG
AHCTDAVNQYTCVCEGYSGLFCEFSPPMVLPTSPCDNFCQNGAQCIIRVNEPICQL
PGYLGEKCEKLVSVNFVNKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDQDHI
ELYRGRVRASYDTGSHPASAIYSVETINDGNFHVIVELLTLDSSLSLSDGGSFKIITNLS
KQSTLNFDSPLYVGGMPGKNNVASLRQAPGQNGTSFHHGCIIRNLYINSELQDFRKVPMQTG
25 ILPGCEPCHKVKCAHGTCQPSQSGFTCECEEGWMGPLCDQRTNDPCLGNKCVHGTCLPI
NAPYSYCKCLEGGHGVLCDEEEDLFNFCQVIKCKHKGKRLSGLGQPYCECSSGFTGDS
REISCRGERIRDYQKQGGYAAOQTTKVSRLECRGGCAGGQCCGPLRSKRRKYSFECTD
GSSFVDEVEKVKCGCTRCAS

```

## SEQ ID NO: 235

30 gi|23238206|ref|NM\_014452.3| Homo sapiens tumor necrosis factor receptor superfamily, member 21 (TNFRSF21), mRNA

```

1 gccaccacgt gtgtccctgc gcccggtggc caccgactca gtccctcgcc gaccagtctg
61 ggcagcggag gagggtggtt ggcagtggtt ggaagcttcg ctatgggaag ttgttccttt
121 gctctctcgc gccagtcct cctccctggt tctcctcagc cgctgtcgga ggagagcacc
35 181 cggagacgcg ggctgcagtc cgcggcgctt ctccccgcct gggcggcgcc gccgctgggc
241 aggtgctgag cgcccctaga gctcccttg ccgcctccct cctctgcccg gccgcagcag
301 tgcacatggg gtgttgagg tagatgggt cccggcccg gaggcggcg tggatgggc
361 gctgggcaga agcagccgc gattccagct gcccgcgcg ccccgggcg cctgcgagt
421 ccccggttca gccatgggga cctctccgag cagcagcacc gccctcgct cctgcagcg
40 481 catcgccgc cgagccacag ccacgatgat cgcgggctcc cttctcctgc ttggattcct
541 tagcaccacc acagctcagc cagaacagaa ggctcgaat ctcatggca cataccgcca
601 tgttgaccgt gccaccggcc aggtgctaac ctgtgacaag tgtccagcag gaacctatgt
661 ctctgagcat tgtaccaaca caagcctgcg cgtctgcagc agttgccctg tggggacctt
721 taccaggcat gagaatggca tagagaaatg ccatgactgt agtcagccat gcccatggcc
45 781 aatgattgag aaattacctt gtgctgcctt gactgaccga gaatgcactt gccacctgg
841 catgttccag tctaacgcta cctgtgcccc ccatacggtg tgtcctgttg gttgggggtg
901 gcggaagaaa gggacagaga ctgaggatgt gcggtgtaag cagtgtgtc ggggtacctt
961 ctcagatgtg cttctagtg tgatgaaatg caaagcatac acagactgtc tgagtccaga
1021 cctgggtggtg atcaagccgg ggaccaagga gacagacaac gtctgtggca cactccgctc
50 1081 cttctccagc tccacctcac cttcccttgg cacagccatc tttccagcc ctgagcacat
1141 ggaaacccat gaagtccctt cctccactta tgttcccaaa ggcatagaact caacagaatc
1201 caactcttct gcctctgtta gaccaaaggt actgagtagc atccaggaag ggacagtcctc
1261 tgacaacaca agctcagcaa gggggaagga agacgtgaac aagaccctcc caaaccttca
1321 ggtagtcaac caccagcaag gccccacca cagacacatc ctgaagctgc tgccgtccat

```

1381 ggaggccact gggggcgaga agtccagcac gcccatacaag ggccccaaga ggggacatcc  
 1441 tagacagaac ctacacaagc attttgacat caatgagcat ttgccctgga tgattgtgct  
 1501 tttcctgctg ctgggtgctg tggtgattgt ggtgtgcagt atccggaaaa gctcgaggac  
 5 1561 tctgaaaaag gggcccggc aggatccag tgccattgtg gaaaaggcag ggctgaagaa  
 1621 atccatgact ccaaccaga accgggagaa atggatctac tactgcaatg gccatggtat  
 1681 cgatatcctg aagctttagt cagcccaagt gggaagccag tggaaagata tctatcagtt  
 1741 tctttgcaat gccagtgaag gggaggttgc tgctttctcc aatgggtaca cagccgacca  
 1801 cgagcgggccc tacgcagctc tgcagcactg gaccatccgg ggccccgagg ccagcctcgc  
 1861 ccagctaatt agcgccttgc gccagcaccg gagaaacgat gttgtggaga agattcgtgg  
 10 1921 gctgatggaa gacaccaccc agctggaaac tgacaaacta gctctcccga tgagccccag  
 1981 cccgcttagc ccgagcccca tccccagccc caacgcgaaa cttgagaatt ccgctctcct  
 2041 gacggtggag ccttcccac aggacaagaa caagggttc ttcgtggatg agtcggagcc  
 2101 ccttctccgc tgtgactcta catccagcgg ctctcccgcg ctgagcagga acggttcctt  
 2161 tattaccaa gaaaagaagg acacagtgtt gcggcaggta cgctggacc cctgtgactt  
 15 2221 gcagcctatc tttgatgaca tgctccactt tctaaatcct gaggagctgc ggggtgattga  
 2281 agagattccc caggctgagg acaactaga ccggtattc gaaattattg gagtcaagag  
 2341 ccaggaagcc agccagaccc tcttgactc tgtttatagc catcttctcct acctgctgta  
 2401 gaacataggg atactgcatt ctggaaatta ctcaatttag tggcaggggt gttttttaat  
 2461 tttctctctg tcttgatttt tggtgtgtg ggtgtgtgtg tgtgtttgtg tgtgtgtgtg  
 20 2521 tgtgtgtgtg tgtgtgtgtg ttttaacagag aatatggcca gtgcttgagt tcttctcct  
 2581 tctctctctc tctttttttt ttaataaact cttctgggaa gttggtttat aagcctttgc  
 2641 cagggtgaac tgttgtgaaa taccaccac taaagtttt taagttccat attttctcca  
 2701 ttttgcttcc ttatgtattt tcaagattat tctgtgcact ttaaatctac ttaacttacc  
 2761 ataatgcag tgtgactttt cccacacact ggattgtgag gctcttaact tcttaaaagt  
 25 2821 ataattggcat cttgtgaatc ctataagcag tctttatgtc tcttaacatt cacacctact  
 2881 ttttaaaaac aaatattatt actattttta ttattgtttg tctttataa attttcttaa  
 2941 agattaagaa aatttaagac cccattgagt tactgtaatg caattcaact ttgagttatc  
 3001 ttttaaatat gtcttgata gttcatattc atggctgaaa cttgaccaca ctattgctga  
 3061 ttgtatgggt ttcacctgga caccgtgtag aatgcttgat tacttgact cttcttatgc  
 30 3121 taatatgctc tgggctggag aaatgaaac ctcaagccat caggatttgc tatttaagtg  
 3181 gcttgacaac tgggccacca aagaactga acttcacctt ttaggatttg agctgttctg  
 3241 gaacacattg ctgcactttg gaaagtcaa atcaagtgcc agtggcgccc tttccataga  
 3301 gaatttgccc agctttgctt taaaagatgt cttgttttt atatacacat aatcaatagg  
 3361 tccaatctgc tctcaaggcc ttggtcctgt tgggattcct tcaccaatta ctttaattaa  
 35 3421 aaatggctgc aactgtaaga acccttgtct gatataattg caactatgct cccatttaca  
 3481 aatgtacctt ctaatgtca gttgccagg tccaatgcaa aggtggcgtg gactcccttt  
 3541 gtgtgggtgg ggtttgtggg tagtggtgaa ggaccgatat cagaaaaatg ctttcaagtg  
 3601 tactaattta ttaataaaca ttaggtgttt gttaaaaaaa aaaaaaaaaa aaaaaaaaaa  
 3661 aa

40 SEQ ID NO: 236

Amino acid sequence of human TNFRSF21 encoded by the DNA sequence shown in SEQ ID NO: 235.

45 MGTSPSSSTALASCSRIARRATATMIAGSLLLLGFLSTTTAQPEQKASNLIGTYRHVDRA  
 TGQVLTCDKCPAGTYVSEHCTNTSLRVCSSCPVGTFRHENGIEKCHDCSQPCPWPMEK  
 LPCAALTDRECTCPGFMFQSNATCAPHTVCPVGWVRKKGTEDEVRCKQCARGTFSDVP  
 SSVMKCKAYTDCLSONLVVVKPGTKETDNVCGTLPSFSSSTSPSPGTAFPRPEHMETHE  
 VPSSTYVPKGMNSTESNSSASVRPKVLSSIQEGTVPDNTSSARGKEDVNKTLPLNLQVNH  
 50 QQGPHRHILKLLPSMEATGGEKSSTPIKPKRGHPRQNLHGHFDINEHLPWMIVLFLLL  
 VLVVIVVCSIRKSSRTLKKGPRQDPSAIVEKAGLKKSMTPTQNREKWIYYCNGHGIDILK  
 LVAAQVGSQWKDIYQFLCNASEREVAAFSNGYTADHERAYALQHWITIRGPEASLAQLIS  
 ALRQHRNDVVEKIRGLMEDTTQLETDKLALPMSPLSPSPIPSPNAKLENSALLTVEP  
 SPQDKNKGFFVDESEPLLRCDSTSSGSSALSRNGSFITKEKDTVLRQVRLDPCDLQFIF  
 DDMLHFLNPEELRVIEEIPQAEKDLRLFEIIGVKSQEASQTLLDSVYSHLPDLL

SEQ ID NO: 237

gi|31341673|ref|NM\_178589.2| Mus musculus tumor necrosis factor receptor superfamily, member 21 (Tnfrsf21), mRNA

```

1 aggtgtcccc gagctgagtg gccatccgac tcagtccctc gccggccgggt ctaggcagcg
61 gaggaggcga gtgcttgata gtggctggaa gcttcgctat gggagtcgc tctttagcac
5 121 tgtcgcggct agccctgctc tctggttctc cgcagccgct gtcgttggag agcaccggga
181 ggcgcgggtt ggcagcgcgc ctgcttctca ccgcccgggc gccagcgcgc ctgggcagggt
241 gctgagcgcc ttccggagcc tcccctgctg ggaggtagat gggctcaccg cccgtgaggc ggcggtggat
301 tgcagtgcac atgggctgct ggaggtagat gggctcaccg cccgtgaggc ggggcccagc gccccgggcg
361 gcggcgctgg gcagaaacag ccaccaattc cagctgccgt ggggcccagc gccccgggcg
10 421 ccgctgcgag ccccgagctc ggccatgggg acccgggcaa gcagcatcac cgcctcgcc
481 tcttgagacc gcaccgccgg ccaagtcgga gccacgatgg tcgcccggctc tcttctctg
541 cttggattcc tcagcaccat cacagctcaa ccagaacaaa agactctgag tctccctggc
601 acctaccgcc atgttgaccg taccactggc caggtgctaa cctgcgacaa gtgcccagca
661 ggaacgtacg tctccgagca ctgtaccaac atgagcctgc gactctgcag cagctgcccc
15 721 gcggggaccc ttaccaggca cgagtgccct atagagagat gccatgactg tagtcagcca
781 tgtccatggc cgatgattga gagattacct tgtgctgcct tgactgaccg agactgacac
841 tgcccacctg gaatgtatca gtctaattggt acctgcgctc ccatacagt gtgcccctg
901 ggctggggtg tgcggaagaa agggacagag aatgaagatg tgcgctgtaa gcagtgcgct
961 cggggtagct tctctgagct gccttccagt gtgatgaagt gtaaagctca caccgactgt
20 1021 ctgggtcaga acctggaggt ggtcaagcca gggaccaagg agacagacaa cgtctgtggc
1081 atgcgcctgt tcttctccag cacaacccca ccttctctg gcacagttac cttttctcac
1141 cctgagcata tggaaatcca cgatgtccct tcctccacct atgagcccca aggcataaac
1201 tcaacagatt ccaactctac tgctctgtgt agaactaagg tacciaagtg catcgaggaa
1261 gggacagtgc ctgacaatac gagctcaacc agtgggaagg aaggcactaa taggacctg
25 1321 ccaaacccac cacaagttac ccaccagcaa gccccccacc acagacacat tctgaagctg
1381 ctgccatcgt ccatggaggc caccgggtgag aagtccagca cagccatcaa ggcccccaag
1441 aggggtcacc ccagacagaa cgctcacaa gatttcgaca tcaacgagca cttgccttgg
1501 atgatcgctc tcttctctct gctggctcctg gtgctgatag tgggtgtgag tatccgaaag
1561 agctccagga ctctcaaaaa ggggcccggg caggatccca gcgccatagt ggaaaggcgc
30 1621 gggctgaaga agtccctgac tcccaccagc aaccgggaga aatggatcta ctaccgcaac
1681 ggccatggta ttgacatctt gaagcttgta gcagcccagg tgggaagcca gtggaaggac
1741 atctatcagt ttctttgcaa cgccagcgag agggaggtgg cggccttctc caatggatac
1801 actgcagatc atgaacgggc ctacgcggct ctgcagcact ggaccatccg tggccctgag
1861 gccagccttg cccagctcat tagcgccttg cgccagcacc gacgcaatga tgtgtggag
35 1921 aagattcggt ggctgatgga agacaccacg cagttggaaa cagacaaact ggctctcccc
1981 atgagcccca gtccgcttag cccgagcccc atgcccagtc ctaacgtgaa acttgagaat
2041 tccactctcc tgacagtggg gccctcaccc ctggacaaga acaagtgtct cttcgtggac
2101 gactcagagc cccttctgag atgcgactcc acatccagtg gctcttcagc actgagcaga
2161 aacggctcct ttattaccaa agaaaagaag gacacagtgt tgcggcaggc ccgctggac
40 2221 cctgtgact tgacgcccac ctttgatgac atgctgcata tcctgaaccc cgaggagctg
2281 cgggtgattg aagagattcc ccaggctgag gacaaactgg accgcctctt cgagatcatt
2341 ggggtcaaga gccaagaagc cagccagacc ctcttgagct ctgtgtacag tcatcttctt
2401 gacctattgt agaacacagg ggcactgcat tctgggaatc aacctactgg cggggtgatt
2461 tcatttcggt tctgactttt gtgttttggg gtgtatgtat gtgtttaaca gactgtatgg
45 2521 ccggtgagtt tgggttcttt ctttctttct ttctttcttt ctttctttct ttctttcttt
2581 ctttctttct ttctttcttt cttccttctt gaaagtgaat gtataaagcc tttaaatgt
2641 ataactgttg gaaaatgccc accactaaat ttttttttaa gttccacata ttctccattt
2701 ttgccttctt atatatatct tcaacactat tctgtgcact ttaaaaactt aacataaacg
50 2761 cagtgtgact tctcccatat gctgggtccc gagactctca acttcttaaa aacctaatgg
2821 catcttgtga ctctagaagc tagacataag tctttcaacc ttcacaccta ctctttctgt
2881 ttttaattatt attgtctatt gtcttttggg ttgtgcttta caagcgttct tgaggacgga
2941 gggatcttac gaccctgttg atgactgtaa ctctattcga ctttgagttg tcttcttcat
3001 gtcttggtat atagttcata ttcatggctg aaacttgacc atactcccta gcgctgattg
3061 tatgggtttc gtctggacac cgtacactgc ctgataactt gtgcacctct taacgctact
55 3121 atgctctggg ctggagaatg aaatctttta gtaccaggga cttgctgttt cagtggcttg
3181 acacctgggc caccaaagaa ctcgatcttc atcttttagg gacacctctg ctgcaccttg
3241 gaaagccaac ctttaagtgc agtggcactt tatgcccagc tttgctttga aagatatctt
3301 tctgtgtttt tttatctctc tctttctctc ttttttttaa aaatacacat agtcaatagg
3361 tccagtctgc cctcaaggcc ttgctgggtt ttcttcatca tccaatcact ttcattaaaa

```

3421 atggetgcag ctgtaagaac tcttgtctga taaattttca actatgctct catttatcta  
 3481 cctgccctct gatgctcagt cgctcagactc taatgcaaag gtggacgtcg gctgcctttg  
 3541 cgtgggcggg cttagtgggt aggaactgat atcagaaaaa aaatgccttc aagtatacta  
 3601 atttattaat aaatattagg tgtttgtt

## 5 SEQ ID NO: 238

Amino acid sequence of mouse TNFRSF21 encoded by the DNA sequence shown in SEQ ID NO: 237.

10 MGTRASSITALASCRTAGQVGATMVAGSLLLLGFLSTITAQPEQKTLSPGTYRHVDRT  
 TGQVLTCDKCPAGTYVSEHCTNMSLRVCSSCPAGTFTRHENGIERCHDCSQPCPWPPIER  
 LPCAALTDRECICPPGMYQSNGTCPHTVCPVGVGVRKKG TENEDVRCKQCARGTFSDVP  
 SSVMKCKAHTDCLGQNLVVKPGTKETDNVCGMRLFFSSTNPPSSGTVTF SHPEHMESH  
 VPSSTYEPQGMNSTDNSNSTASVRTKVP SGIEEGTVPDNTSSTSGKEGTNRTL PNP PQVTH  
 15 QQAPHRRHILKLLPSSMEATGEKSSTA IKA PKRGHPRONAHKHFDINEHLPWMIVLFLLL  
 VLVLIIVVCSIRKSSRTLKKGPRQDPSAIVEKAGLKKSLTPTQNREKWIYYRNGHGIDILK  
 LVAAQVGSQWKDIYQFLCNASEREVA AFSNGYTADHERAYAA LQHW TIRGPEASLAQLIS  
 ALRQHRNDVVEKIRGLMEDTTQLETDKLALPMSPLSPSPMPSPNVKLENSTLLTVEP  
 SPLDKNCKFFVDESEPLLRCDSTSSGSSALSRNGSFITKEKKDTVLRQVRLDPDCLQPIF  
 DDMLHLNPEELRVIEEIPQAEDKLDRLFEIIGVKSQEASQTLLDSVYSHLPDLL

## SEQ ID NO: 239

20 gi|34874517|ref|XM\_236992.2| Rattus norvegicus similar to death receptor 6 (LOC316256), mRNA

1 ccgggaggcg cgggttgcaa gcgcgcctgc ttctccccgc ccggggcgcca gcgcccctgg  
 61 gcagggtgctg agegccttct ccgagcctcc cctgctgcct ccctcttccg cctagccgcc  
 121 ttgctgctgc agtgacatg ggctgctgga ggtagatggg ctcaccgccc gtgaggcggc  
 25 181 ggtggatgcg gcgctgggca gaaacagcca ccgattccag ctgccgtggg gccgagcgcc  
 241 cggagcgccg ctgcgagccc cgggctcagc catggggacc tccgcaagca gcatcacgcg  
 301 cctegcctct tgcagccgca tgcggcgcca agttggagcc acgatggctg ccggctccct  
 361 tctcttgctt gggttcctca gcaccatcac agcccaacca gaacaaaaga ctctgagtct  
 421 caccgggcacg tacgcgccag ttgaccgtac cactggccag gtgctaacct gcgacaagtg  
 30 481 tccggcagga acgtatgtct ccgagcactg taccaacacg agcctgcgag tctgcagcag  
 541 ctgcccctcg gggaccttta ccaggcatga gaacggcata gagagatgcc atgactgtag  
 601 tcagccatgc ccacggccga tgattgagag attaccttgt gctgccttga ctgaccgaga  
 661 atgcatctgc ccacccggaa tgatcagtc taatgggacc tgcgctcccc acacggtgtg  
 721 ccccggtgggc tggggtgtga ggaagaaagg gacagagaat gaagatgtgc ggtgtaagca  
 35 781 gtgtgctcgc ggtaccttct ctgacgtgcc ttccagtgtg atgaagtgtg gagccacac  
 841 ggactgtctg ggtcagaacc tgatggtggg caagcagggg actaaggaga cagacaacgt  
 901 ctgtggcgctg cacctgtcct cctccagcac gaccccatct tcccctggca tagctacctt  
 961 ttctcatcct gagcacacgg aatcccacga tgtcccctcc tccacctatg agccccaagg  
 1021 catgaactca acagattcca actctactgc ctctgttaga actaaggtag caagtacat  
 40 1081 ccaggaagag acagtgcctg acaatacaag ctccacgagt gggaaggaga gcacgaacag  
 1141 gaccctgcc aacccaccac aacttaccac ccagcaaggc cccaccaca gacacattct  
 1201 gaagctgctg ccatcatcca tggaggccac tggtgagaag tccagcacag ccatcaaggc  
 1261 ccccaagagg ggcattccca acacaagcat ttcgacatca acgagcattt  
 1321 gccctggatg attgtactct tccttctgct ggtcctgggt ctcatgtgtg tgtgcagtat  
 45 1381 ccgaaagagc tctaggactc tcaagaagg gccccggcag gatcccgatg ccatcatgga  
 1441 aaaggcaggg ctgaagaagt ccttgactcc aaccagaac cgggagaaat ggatctacta  
 1501 ccgcaatggc cacgggtattg acatcctgaa gcttgtagca gcgcaggtgg gaagccagtg  
 1561 gaaggacatc tatcagtttc tttgcaacgc cagcgagagg gaggtggctg ccttctccaa  
 1621 tggatacact gcagaccatg aacgagccta tgcagctctg cagcactgga ccatccgggg  
 50 1681 ccttgaggcc agccttgctc agctcatcag cgccttgctc cagcaccgac gcaatgatgt  
 1741 tgtggagaag attcgtgggc tgatggaaga tactacgcag ttggaaaccg acaaatggc  
 1801 tctccccatg agccccagtc cgcttagccc gagccccatt cccagtccta acgtgaaact



1861 tgagaattcc actctcctga cagtggagcc ctcacccctg gacaagaaca agggcttctt  
 1921 cgtggacgag tcggagcccc ttctgcgttg tgactccacg tccagcggtt cttcagcact  
 1981 gagcaggaat ggctccttta taaccaaaaga aaagaaggac acagtgttgc ggcagggtccg  
 2041 cctggacccc tgtgacttgc agcccatctt tgacgacatg ctgcacatcc tgaatccgga  
 5 2101 agagctgcgg gtaatcgaag agattcccca ggctgaggac aaactggacc gtctcttcga  
 2161 gatcattggg gtcaagagcc aagaagccag ccagaccctc ttggactctg tgtacagcca  
 2221 tcttctcgac ctattgtag

SEQ ID NO: 240

10 Amino acid sequence of rat TNFRSF21 encoded by the DNA sequence shown in SEQ ID NO: 239.

MGTSASSITALASCRIAGQVGATMVAGSLLLLGFLSTITAQPEQKTLSTGTyrHVDRT  
 TGQVLTCDKCPAGTYVSEHCTNTSLRVCSSCPSTFTRHENGIERCHDCSQPCPRMIER  
 LPCAALTDRECICPPGMYQSNGTCAPTVCPVGGWVRKKGTENEDVRCKQCARGTFSDVP  
 15 SSVMKCRAHTDCLGQNLMMVVKQGTKETDNVCGVHLSSSSSTPSSPGIATFSHPHTESH  
 VPSSTYEPQGMNSTDSNSTASVRTKVPSDIQEETVPDNTSSTSGKESTNRTLPPNPQLTH  
 QQGPHHRHILKLLPSSMEATGEKSSTAIKAPKRGHPRQNPCHKHFDINEHLPWMI VLFLL  
 VLVLIVVCSIRKSSRTLKKGPRQDPSAIMEKAGLKKSLTPTQNRKWIYYRNGHGIDILK  
 LVAAQVGSQWKDIYQFLCNASEREVAAFSNGYTADHERAYAALQHWITIRGPEASLAQLIS  
 20 ALRQHRNDVVEKIRGLMEDTTQLETDKLALPMSPLSPSPISPNNVKLENSTLLTVEP  
 SPLDKNKGFFVDESEPLLRCDSTSSGSSALSRNGSFITKEKDTVLRQVRLDPCDLQPIF  
 DDMHLHILNPEELRVIREIPQAEDKLDRLFETIGVKSQEQASQTLLDSVYSHLPDLL

SEQ ID NO: 241

gi|23510443|ref|NM\_006573.3| Homo sapiens tumor necrosis factor (ligand) superfamily, member 13b (TNFSF13B), mRNA

25 1 gaaattctta caaaaactga aagtgaatg aggaagacag attgagcaat ccaatcggag  
 61 ggtaaatgcc agcaaaccta ctgtacagta ggggtagaga tgcagaaagg cagaaaggag  
 121 aaaattcagg ataactctcc tgaggggtga gccaaagccct gccatgtagt gcacgcagga  
 181 catcaacaaa cacagataac aggaatgat ccattccctg tggtcactta ttctaaaggc  
 241 cccaaccttc aaagttcaag tagtgatatg gatgactcca cagaaaggga gcagtcacgc  
 30 301 cttacttctt gccttaagaa aagagaagaa atgaaactga aggagtgtgt ttccatctc  
 361 ccacggaagg aaagcccctc tgtccgatcc tccaaagacg gaaagctgct ggctgcaacc  
 421 ttgctgctgg cactgctgtc ttgctgcctc acggtggtgt ctttctacca ggtggccgcc  
 481 ctgcaagggg acctggccag cctccgggca gagctgcagg gccaccacgc ggagaagctg  
 541 ccagcaggag caggagcccc caaggccggc ctggaggaag ctccagctgt caccgcggga  
 35 601 ctgaaaatct ttgaaccacc agctccagga gaaggcaact ccagtcagaa cagcagaaat  
 661 aagcgtgccg ttcagggtcc agaagaaaca gtcactcaag actgcttgca actgattgca  
 721 gacagtgaac caccaactat acaaaaagga tcttacacat ttgttccatg gcttctcagc  
 781 tttaaaaggg gaagtgcctc agaagaaaaa gagaataaaa tatttggtcaa agaaactggt  
 841 tactttttta tatatggtca ggttttatat actgataaga cctacgccat gggacatcta  
 40 901 attcagagga agaaggtcca tgtctttggg gatgaattga gtctggtgac tttgtttcga  
 961 tgtattcaaa atatgcctga aacactaccc aataattcct gctattcagc tggcattgca  
 1021 aaactggaag aaggagatga actccaactc gcaataccaa gagaaaatgc acaaatatca  
 1081 ctggatggag atgtcacatt ttttggtgca ttgaaactgc tgtgacctac ttacaccatg  
 1141 tctgtagcta ttttctctcc tttctctgta cctctaagaa gaaagaatct aactgaaaat  
 45 1201 acca

SEQ ID NO: 242

Amino acid sequence of human TNFSF13B encoded by the DNA sequence shown in SEQ ID NO: 241.

MDDSTEREQSRLTSLCKKREEMKLKECVSILPRKESPSVRSSKDGKLLAATLLLALLSCC  
 LTVVSFYQVAALQGD LASLRAELQGHHA EKLPAGAGAPKAGLEEAPAVTAGLKIFEP PAP  
 GEGNSSQNSRNKRAVQGP EETVTQDCLQLIADSETPTIQG SYTFVPWLLSFKRGSAL EEEKENKILVKETGYF  
 KENKILVKETGYFFIYGQVLYTDKTYAMGH LIQRKKVHVFGDELSLVTLFR CIQNMPETL  
 5 PNNSCYSAGIAKLEEGDELQLAIPRENAQISLDG DVTFFGALKLL

SEQ ID NO: 243

Amino acid sequence of human TNFSF13B, a soluble active secreted form derived from SEQ ID NO:242.

10 AVQGP EETVTQDCLQLIADSETPTIQG SYTFVPWLLSFKRGSAL EEEKENKILVKETGYF  
 FIYGQVLYTDKTYAMGH LIQRKKVHVFGDELSLVTLFR CIQNMPETL PNNSCYSAGIAKL  
 EEGDELQLAIPRENAQISLDG DVTFFGALKLL

SEQ ID NO: 244

gi|32454911|gb|AY302751.1| Homo sapiens delta BAFF (TNFSF13b) mRNA, complete cds;  
 alternatively spliced

15 1 atggatgact ccacagaaag ggagcagtc a cgccttactt cttgccttaa gaaaagagaa  
 61 gaaatgaaac tgaaggagtg tgtttccatc ctcccacgga aggaaagccc ctctgtccga  
 121 tcttccaaag acggaaagct gctggctgca accttgctgc tggcactgct gtcttgctgc  
 181 ctcacgggtg tgcttttcta ccagggtggc gccctgcaag gggacctggc cagcctccgg  
 241 gcagagctgc agggccacca cggcgagaag ctgccagcag gagcaggagc cccaaggcc  
 20 301 ggctggagg aagctccagc tgcaccgcg ggactgaaaa tctttgaacc accagctcca  
 361 ggagaaggca actccagtc a gaacagcaga aataagcgtg ccgttcaggg tccagaagaa  
 421 acaggatctt acacatttgt tccatggctt ctcagcttta aaaggggaag tgccctagaa  
 481 gaaaaagaga ataaaatatt ggtcaaagaa actggttact tttttatata tggtcagggt  
 541 ttatatactg ataagacct a cgccatggga catctaattc agaggaagaa ggtccatgtc  
 25 601 tttggggatg aattgagctt ggtgactttg tttcgatgta ttcaaaatat gcctgaaaca  
 661 ctacccaata attcctgcta ttcagctggc attgcaaaac tggaagaagg agatgaactc  
 721 caacttgcaa taccaagaga aaatgcacaa atatcactgg atggagatgt cacatttttt  
 781 ggtgcattga aactgctgtg a

SEQ ID NO: 245

30 Amino acid sequence of human TNFSF13B variant ORF number 1 encoded by the DNA  
 sequence shown in SEQ ID NO: 244.

MDDSTEREQSRLTSLCKKREEMKLKECVSILPRKESPSVRSSKDGKLLAATLLLALLSCC  
 LTVVSFYQVAALQGD LASLRAELQGHHA EKLPAGAGAPKAGLEEAPAVTAGLKIFEP PAP  
 GEGNSSQNSRNKRAVQGP EETG SYTFVPWLLSFKRGSAL EEEKENKILVKETGYFFIYGQV  
 35 LYTDKTYAMGH LIQRKKVHVFGDELSLVTLFR CIQNMPETL PNNSCYSAGIAKLEEGDEL  
 QLAI PRENAQISLDG DVTFFGALKLL

SEQ ID NO: 246

gi|32441946|gb|AY290823.1| Mus musculus delta BAFF (Tnfsf13b) mRNA, complete cds;  
 alternatively spliced

40 1 atggatgagt ctgcaaagac cctgccacca ccgtgcctct gtttttgetc cgagaaagga  
 61 gaagatatga aagtgggata tgatcccatc actccgcaga aggaggaggg tgcttggttt  
 121 gggatctgca gggatggaag gctgctggct gctaccctcc tgetggccct gttgtccagc  
 181 agtttcacag cgatgtcctt gtaccagttg gctgccttgc aagcagacct gatgaacctg  
 241 cgcagtgagc tgcagagcta ccgaggttca gcaacaccag ccgccgcggg tgctccagag

301 ttgaccgctg gagtcaaact cctgacaccg gcagctcctc gaccccacaa ctccagccgc  
 361 ggccacagga acagacgcgc ttccaggga ccagaggaaa cagaacaaga tgtagacctc  
 421 tcagctcctc ctgcaccatg cctgcctgga tgccgccatt ctcaacatga tgataatgga  
 481 atgaacctca gaaacagaac ttacacattt gttccatggc ttctcagctt taaaagagga  
 5 541 aatgccttgg aggagaaaga gaacaaaata gtggtgaggg aaacaggcta tttcttcac  
 601 tacagccagg ttctatacac ggaccccatc ttgctatgg gtcagtcat ccagagggaag  
 661 aaagtacacg tctttgggga cgagctgagc ctggtgaccc tggtccgatg tattcagaat  
 721 atgccccaaa cactgcccac caattcctgc tactcggctg gcacgcgag gctggaagaa  
 781 ggagatgaga ttcagcttgc aattcctcgg gagaatgcac agatttcacg caacggagac  
 10 841 gacaccttct ttggtgcctt aaaactgctg taa

SEQ ID NO: 247

Amino acid sequence of mouse TNFSF13B encoded by the DNA sequence shown in SEQ ID NO: 246.

15 MDESAKTLPPPCLCFCSEKGEDMKVGYDPI TPQKEGAWFGICRDGRLLAATLLLLALLSS  
 SFTAMSLYQLAALQADLMNLRMELQSYRGSATPAAAGAPELTAGVKLLTPAAPRPHNSSR  
 GHRNRRAFQGPETEEDVDLSAPPAPCLPGCRHSQHDDNGMNLNRNTYTFVPWLLSFKRG  
 NALBEEKENKIVVRQTGYFFIYSQVLYTDFIFAMGHVIRKKVHVFGEDELSLVTLFRICIQN  
 MPKTLPNNSCYSAGIARLEEGDEIQLAIPRENAQISRNGDDTFFGALKLL

SEQ ID NO: 248

20 ENSRNOT00000019397 cDNA sequence, EnsEMBL transcript [Rattus norvegicus]

1 atggataagt ctgcaaagaa cctgccacca ccgcgcctct gtttttgccc tgagaaagga  
 61 gaagatatga aagtgggata tggtcccatc actccgcaga aggagggtgc ctgggttggg  
 121 atctgcaggg acagacggct gctggtgct accctcctgc tggctctgtt gtccagcagt  
 181 ttcacagcga tgtccttgta ccagttggct gtcttgcaag cagacctgat gagcctgcgc  
 241 atggagctgc agagctaccg aagttcacgc acaccgcgc ccccgggggc tccagggttg  
 25 301 tccgcggggag tcaaaactccc aacaccgcga gctcctggac cccacaactc cagccgaggc  
 361 caaaggaaca gacgcgcttt ccagggaccg gaggaacag aacaagatgt agacctctca  
 421 gctactccag tgccatccct gcctggaac tgccatgctt ctcaccatga tgagaatgga  
 481 ctgaacctca gaaccatcat tcaagactgt ctgcagctga ttgcagacag caacacgccg  
 30 541 actatacgaa aaggaactta cacatttgtt ccatggcttc tcagctttaa aagaggaaac  
 601 gccttggagg agaaagagaa caaaatagtg gtgaggcaaa caggctatctt cttcatctac  
 661 agccaggttc tgtacacgga ccccatctt gccatgggtc atgtcatcca gaggaagaaa  
 721 atacacgtgt ttggggatga gctgagcctg gtgactctgt tccgatgcat ccagaatatg  
 781 ccgaaaacac tgcccaacaa ttctgtctac tccgctggca tcgcgaagct ggaagaaggg  
 35 841 gacgagattc agcttgcaat acctcgggag aatgcccgaga ttccacggaa cggagacgac  
 901 accttctttg gtgccctaaa actgctgtga

SEQ ID NO: 249

Amino acid sequence of rat TNFSF13B encoded by the DNA sequence shown in SEQ ID NO: 248.

40 MDKSAKNLPPRLCFCPEKGEDMKVGYVPI TPQKEGAWVGICDRRLAATLLLLALLSSS  
 FTAMSLYQLAVLQADLMNLRMELQSYRSSATPAPGAPGLSAGVKLPTPAAPGPHNSSRG  
 QRNRRAFQGPETEEDVDLSATVPVSLPGNCHASHHDENGLNLRITIIQDCLQLIADSNTP  
 TIRKGTYYTFVPWLLSFKRGNALBEEKENKIVVRQTGYFFIYSQVLYTDFIFAMGHVIRKK  
 45 IHVFGEDELSLVTLFRICIQNPMTLPNNNSCYSAGIAKLEEGDEIQLAIPRENAQISRNGDD  
 TFFGALKLL

SEQ ID NO: 250

gi|25952143|ref|NM\_003807.2| Homo sapiens tumor necrosis factor (ligand) superfamily, member 14 (TNFSF14), transcript variant 1, mRNA

```

1  cgagactcca tctcaaaaac aaaacaaata aacgaacaaa aaaaccacac acgtattatt
5  61  ttcttggtta cgaggtttct tgtctctctg gctccaccag aagaggagca gggacccttc
121 ttgctgttgt tcattgctgc atccccaca ccgagagcag agcctggcat gggcagaaag
181 tcctcagtcg atatttggtg gcccacagcg aatgaagcat ccaagaaggg aaagctgggg
241 gctccccact gcacttgcca cctgagtcac attttcagaa gcctctggaa agtcgtgcac
301 agcccaggag tggtgagcaa ttctgggttc ctctgagggt gaaggacca ggcgtgtcag
361 ccttgctcca gacaccttgg gcatggagga gagtgtcgta cggccctcag tgtttgggt
10 421 ggatggacag accgacatcc cattcacgag gctgggacga agccaccgga gacagtcgtg
481 cagtgtggcc cgggtgggtc tgggtctctt gctgttctg atgggggccc ggctggccgt
541 ccaaggctgg ttcctcctgc agctgcactg gcgtctagga gagatggtca cccgctgccc
601 tgacggacct gcaggctcct gggagcagct gatacaagag cgaaggcttc acgagggtcaa
661 cccagcagcg catctcacag gggccaactc cagcttgacc ggcagcgggg ggccgctggt
15 721 atgggagact cagctgggcc tggccttctt gaggggcctc agctaccacg atggggccct
781 tgtggtcacc aaagctggct actactacat ctactccaag gtgcagctgg gcggtgtggg
841 ctgcccgtcg ggcctggcca gcaccatcac ccacggcctc tacaagcgca caccctgcta
901 ccccgaggag ctggagctgt tggtcagcca gcagtcaccc tgcggacggg ccaccagcag
961 ctcccgggtc tgggtgggaca gcagcttctt ggggtggtgt gtacacctgg aggctgggga
20 1021 gaagggtggt gtccgtgtgc tggatgaacg cctgggttga ctgctgtatg gtacctgggtc
1081 ttacttcggg gctttcatgg tgtgaaggaa ggagcgtggt gcattggaca tgggtctgac
1141 acgtggagaa ctccagaggt gctcagggg aaagaaaact cacgaagcag aggctgggcg
1201 tgggtggctc cgctgtgaat cccagcactt tgggaggcca aggcaggcgg atcacctgag
1261 gtcaggagtt cgagaccagc ctggctaaca tggcaaaacc ccatctctac taaaaatata
25 1321 aaaattagcc ggacgtggtg gtgctgcct gtaatccagc tactcaggag gctgaggcag
1381 gataattttg cttaaaccgg ggaggcggag gttgcagtga gccgagatca caccactgca
1441 ctccaacctg ggaaacgcag tgagactgtg cctcaaaaaa aaaaaaaaaa a

```

SEQ ID NO: 251

30 Amino acid sequence of human TNFSF14 encoded by the DNA sequence shown in SEQ ID NO: 250.

```

MEESVVRPSVFVVDGQTDIPFTRLGRSHRRQSCSVARVGLGLLLLLMGAGLAVQGWFLQ
LHWRLGEMVTRLDPGPGASWEQLIQERRSHEVNPAHLTGANSSLTGSGGPLLWETQLGL
AFLRGLSYHDGALVVTKAGYIIYSKVQLGGVGCPLGLASTITHGLYKRTPRYPEELELL
35 VSQQSPCGRATSSSRVWWDSSFLGGVVHLEAGEKVVRVLDRLVRLRDGTRSYFGAFMV

```

SEQ ID NO: 252

Amino acid sequence of human TNFSF14, a soluble active secreted form derived from SEQ ID NO: 251.

```

QERRSHEVNPAHLTGANSSLTGSGGPLLWETQLGLAFLRGLSYHDGALVVTKAGYIIY
40 SKVQLGGVGCPLGLASTITHGLYKRTPRYPEELELLVSQQSPCGRATSSSRVWWDSSFLG
GVVHLEAGEKVVRVLDRLVRLRDGTRSYFGAFMV

```

SEQ ID NO: 253

gi|17390119|gb|BC018058.1| Homo sapiens tumor necrosis factor (ligand) superfamily, member 14, mRNA (cDNA clone MGC:26477 IMAGE:4793299), complete cds

```

45 1  ggtgcacagc ccaggagtgt tgagcaattt cggtttcttc tgagggttaa ggaccagggc
61  gtgtcagccc tgctccagac accttgggca tggaggagag tgctgtacgg ccctcagtgt

```

```

121 ttgtgggtgga tggacagacc gacatcccat tcacgaggct gggacgaagc caccggagac
181 agtcgtgcag tgtggcccgg gtgggtctgg gtctcttgct gttgctgatg ggggccgggc
241 tggccgtcca aggtgggttc ctctgcagc tgcactggcg tctaggagag atggtcaccc
301 gcctgcctga cggacctgca ggctcctggg agcagctgat acaagagcga aggtctcacg
5 361 aggtcaaccc agcagcgcag ctacacagggg ccaactccag cttgaccggc agcggggggc
421 cgctgttatg ggagactcag ctgggcgtgg ccttcctgag gggcctcagc taccacgatg
481 gggcccttgt ggtcaccaaa gctggctact actacatcta ctccaaggtg cagctgggcg
541 gtgtgggctg cccgctgggc ctggccagca ccatcaccca cggcctctac aagcgcacac
601 cccgctaccc cgaggagctg tagctgttgg tcagccagca gtcaccctgc ggacgggcca
10 661 ccagcagctc ccgggtctgg tgggacagca gcttcctggg tgggtgtgga cacctggagg
721 ctggggagga ggtggctcgtc cgtgtgctgg atgaacgcct ggttcgactg cgtgatgga
781 cccggtctta cttcggggct ttcattggtg gaaggaagga gcggtggtga ttggacatgg
841 gtctgacacg tggagaactc agagggtgcc tcaggggaaa gaaaactcac gaagcagagg
901 ctgggcgtgg tggctctcgc ctgtaatccc agcactttgg gaggccaagg caggcggatc
15 961 acctgaggtc aggagttcga gaccagcctg gctaacatgg caaaaccca tctctactaa
1021 aaatacaaaa attagccgga cgtgggtggtg cctgcctgta atccagctac tcaggaggct
1081 gaggcaggat aattttgctt aaaccggga ggcggaggtt gcagttagcc gagatcacac
1141 cactgcactc caacctggga aacgcagtga gactgtgcct caaaaaaaaaa gaaagaaaaa
1201 aaaaaaaaaa a

```

20 SEQ ID NO: 254

Amino acid sequence of human TNFSF14 variant ORF number 1 encoded by the DNA sequence shown in SEQ ID NO: 253.

```

25 MEESVVRPSVFVVDGQTDIPFTRLGRSHRRQSCSVARVGLGLLLLLMGAGLAVQGWFLQ
LHWRLGEMVTRLPGDPAGSWEQLIQRRSHEVNPAHLTGANSSLTSGGPLLWETQLGV
AFLRGLSYHDGALVVTKAGYYYIYSKVQLGGVGCPLGLASTITHGLYKRTPRYPEEL

```

SEQ ID NO: 255

gi|25952146|ref|NM\_172014.1| Homo sapiens tumor necrosis factor (ligand) superfamily, member 14 (TNFSF14), transcript variant 2, mRNA

```

30 1 cgagactcca tctcaaaaac aaaacaaata aacgaacaaa aaaaccacaca acgtattatt
61 ttcttgttta cgaggtttct tgtctctctg gctccaccag aagaggagca gggacccttc
121 ttgctgttgt tcattgctgc atccccaca ccgagagcag agcctggcat gggcagaaaag
181 tcctcagtcg atatttggtg gccccagcg aatgaagcat ccaagaaggg aaagctgggg
241 gctccccact gcaattgcc a cctgagtcac attttcagaa gcctctggaa agtcgtgcac
301 agccccaggag tgttgagcaa tttcggtttc ctctgaggtt gaaggaccca ggcgtgtcag
35 361 cctgctcca gacaccttgg gcatggagga gagtgtcgta cggccctcag tgtttgtggt
421 ggatggacag accgacatcc cattcacgag gctgggacga agccaccgga gacagtctgt
481 cagtgtggcc cgggacggac ctgcaggctc ctgggagcag ctgatacaag agcgaaggctc
541 tcacgaggtc aacccagcag cgcattcac agggggccaac tccagcttga ccggcagcgg
601 ggggcccgtg ttatgggaga ctacagctggg cctggccttc ctgaggggccc tcagctacca
40 661 cgatggggcc cttgtggtca ccaaaagctg ctactactac atctactcca aggtgcagct
721 gggcgggtgt ggctgcccgc tgggcctggc cagcaccatc acccacggcc tctacaagcg
781 cacaccccg c taccggagg agctggagct gttggtcagc cagcagtcac cctgcggacg
841 ggccaccagc agctcccggg tctgggtggga cagcagcttc ctgggtggtg tggtaacact
901 ggaggctggg gaggaggtgg tcgtccgtgt gctggatgaa cgctgggttc gactgcgtga
45 961 tggtagccgg tcttaacttc gggctttcat ggtgtgaagg aaggagcgtg gtgcattgga
1021 catgggtctg acacgtggag aactcagagg gtgcctcagg ggaagaaaaa ctacgaagc
1081 agaggctggg cgtggtggct ctgcctgta atcccagcac tttgggaggc caaggcaggc
1141 ggatcacctg aggtcaggag ttcgagacca gcctggctaa catggcaaaa ccccatctct
1201 actaaaaata caaaaattag ccggacgtgg tgggtgctgc ctgtaatcca gctactcagg
50 1261 aggctgaggc aggataattt tgcttaaac ccggaggcgg aggttgagc gagccgagat
1321 cacaccactg cactccaacc tgggaaacgc agtgagactg tgccctcaaaa aaaaaaaaaa
1381 aaaaaaaaaa aaaaa

```

## SEQ ID NO: 256

Amino acid sequence of human TNFSF14 variant ORF number 2 encoded by the DNA sequence shown in SEQ ID NO: 255.

5 MEESVVRPSVFVVDGQTDIPFTRLGRSHRRQSCSVARDGPAGSWEQLIOERRSHEVNPAA  
HLTGANSSLTGSGGPLLWETQLGLAFLRGLSYHDGALVVTKAGYYYIYSKVQLGGVGCPL  
GLASTITHGLYKRTPRYPELELLVSQQSPCGRATSSSRVWWDSSFLGGVVHLEAGEEVV  
VRVLDRLVRLRDGTRSYFGAFMV

## SEQ ID NO: 257

10 gi|9507194|ref|NM\_019418.1| Mus musculus tumor necrosis factor (ligand) superfamily,  
member 14 (Tnfsf14), mRNA

1 gcccacacag ctccgggcagt ttgcacagcc cgagcgtgtt gggcaattgt ggtttcctcc  
61 ggagaggagg aactcaggct tgccaaccct ttccctgggc ttccggagcct cagctgctct  
121 ggcatggaga gtgtggtaca gccttcagtg tttgtggtgg atggacagac ggacatccca  
181 ttcaggcggc tggaacagaa ccaccggaga cggcgctgtg gcactgtcca ggtcagcctg  
15 241 gccctggtgc tgctgctagg tgctgggctg gccactcagg gctggtttct cctgagactg  
301 catcaacgtc ttggagacat agtagctcat ctgccagatg gaggcaaagg ctccctgggag  
361 aagctgatac aagatcaacg atctcaccag gccaacccag cagcacatct tacaggagcc  
421 aacgccagct tgataggtat tgggtggacct ctgttatggg agacacgact tggcctggcc  
481 ttcttgagg gcttgacgta tcatgatggg gccctggtga ccatggagcc cggttactac  
20 541 tatgtgtact ccaaagtga gctgagcggc gtgggctgcc ccagggggt ggccaatggc  
601 ctccccatca cccatggact atacaagcgc acatcccgc acccgaagga gttagaactg  
661 ctggtcagtc ggcggtcacc ctgtggcccg gccaacagct cccgagtctg gtgggacagc  
721 agcttccttg gcggcgtggt acatctggag gctggggaag aggtggtggt ccgctgcct  
781 ggaaaccgcc tggtcagacc acgtgacggc accaggtcct atttcggagc tttcatggct  
25 841 tgaaggctgc ggtgacaatg tattttgtgg agggacctct ccaggactca cctcaaaccc  
901 agcaataggg tttgaagtcc tccctttaa gaggccctgaa ctctgcagtg ctccggggcg  
961 tgtagactgc tgacctgctt tgggcaatct tcaaatacaga gacctggaga ctggggcgct  
1021 ggagcccagg agcgaggggt cagctcattt gcctgatatt caggaagaaa gaatcaagct  
1081 ggggtattta tgcttctgat gcaaacactg agatttcggc tttctgggtt ttgagctgga  
30 1141 ggcaagaaac cttcccagag tgtcatcagg accatgttgg caggacttgg ggctccagac  
1201 ttgccaccac actctggcct ctcccaccca tccgctgcat tggtttccag ccacaaaaac  
1261 agcactggcc ccctggctgc aactggccag gtacgagctt ctgagcacct acattcctca  
1321 gggacatctt gatgagatct cagtactcag tccaatgcgc agcagcgaca gacatgccag  
1381 gaattggttg tcagaaggga agggaggaag gggaggaag aagggaatgc agaagagaag  
35 1441 gggggaaaac aagacaaaaa caaaacagca acaacaaagc ggcaggagg aggtgacacc  
1501 cttggggata ctttagtcaa cacacttaga acagattgtg ccaggcctgt tggattcctg  
1561 gagttagatg gatcgtggga aggcacaatg gggagcaagt gggcttgggt tatggctcag  
1621 tgggtaaagt gcaattatgg ggatctgagt ttgaatccct ggtaccata taaagacaca  
1681 gatgcggtga tgggcacttg tgacaatgag atcatcaata gggaaatggag acaggaggga  
40 1741 cctctggggt tcaactggcca ggcagtctag ctgaatcaaa gagctccaag ttcagtcgat  
1801 agctcctgaa gatgacaact gaggctattc tccaaacccc acacgcagga cacatgcgta  
1861 at

## SEQ ID NO: 258

45 Amino acid sequence of mouse TNFSF14 encoded by the DNA sequence shown in SEQ ID  
NO: 257.

MESVVQPSVFVVDGQTDIPFRLLEQNHRRCRGTQVQSLALVLLLGAGLATQGWFLRLH  
QRLGDIVAHLDPDGGKGSWEKLIQDQRSHQANPAHLTGANASLIGIGGPLLWETRLGLAF  
LRGLTYHDGALVTMEPGYYYVYSKVQLSGVGCPLGLANGLPITHGLYKRTRYPKELELL  
VSRSPCGRANSSRVWWDSSFLGGVVHLEAGEEVVVRVPGNRLVRPRDGTRSYFGAFMV

## SEQ ID NO: 259

gi|27672737|ref|XM\_236794.1| Rattus norvegicus similar to LIGHT protein (LOC301133), mRNA

```

5      1 atggagagtg tggtagagcc ttccagtgtt gtggtggatg gacagacaga catcccatc
61 aggcggctgg gacagaacca caggagacgg cactgcggca ctgtccaggc cagcctggcc
121 ctgctgctgc tgcctgggtg tgggctggcc actgagggtt ggtttctcct gagactgcat
181 cagcgtcttg gggacatagt agctcatctg ccagatggag gcaaaggctc ctgggagaag
241 ctgatacaag gagctaacgc cagcttgata ggcattggtg gacctctgtt atgggagaca
301 caacttggcc tggccttcct gaggggcctg acgtatcatg atggggccct ggtgaccacc
10    361 gaggtgggtt actactacgt gtactccaaa gtgcagttga gtggtgtggg ctgccccag
421 gggctggcca atggcctccc catcacccac gggctgtaca agcgcacatc ccgatacccc
481 aaggagttag aactgctggt cagccggcgg tcacctgtgt gccgggcca aagctccga
541 gtctggtggg acagtagttt cctcggcgga gtggtacatc tggaggccgg agaagaggtg
601 gtggtccgcg tgcctggaaa ccgcctggtc agaccacgtg atggcacgag gtcctatttc
15    661 ggagctttca tgatctga

```

## SEQ ID NO: 260

Amino acid sequence of rat TNFSF14 encoded by the DNA sequence shown in SEQ ID NO: 259.

```

20    MESVQPSVFVVDGQTDIPFRRLLGQNHRRRHCGTVQVSLALLLLLGAGLATEGWFLRLRH
      QRLGDIVAHLPDGGKGSWEKLIQGANASLIGIGGPLLWETQLGLAFLRGLTYHDGALVTT
      EAGYYVYSKVQLSGVGCPOQLANGLPITHGLYKRTSRYPKELELLVSRSSPCGRANSSR
      VVWDSFLLGGVVHLEAGEVVRVPGNRLVLRPRDGRSYFGAFMI

```

## SEQ ID NO: 261

gi|13775596|ref|NM\_024164.2| Homo sapiens tryptase beta 2 (TPSB2), mRNA

```

25    1 ggccaggatg ctgaatctgc tgotgctggc gctgcccgtc ctggcgagcc gcgcctacgc
61 ggcccctgcc ccaggccagg ccctgcagcg agtgggcacg gttgggggtc aggaggcccc
121 caggagcaag tggccctggc aggtgagcct gagagtccac ggccataact ggatgcaactt
181 ctgcgggggc tccctcatcc acccccagtg ggtgctgacc gcagcgcact gcgtgggacc
241 ggacgtcaag gatctggccg ccctcagggt gcaactgcgg gagcagcacc tctactacca
30    301 ggaccagctg ctgccggtca gcaggatcat cgtgcacca cagttctaca ccgcccagat
361 cggagcggac atcgccctgc tggagctgga ggagccggtg aaggtctcca gccacgtcca
421 cacggtcacc ctgcccctg cctcagagac cttccccccg gggatgccgt gctgggtcac
481 tggctggggc gatgtggaca atgatgagcg cctcccaccg ccatttcttc tgaagcaggt
541 gaaggtcccc ataattgaaa accacatttg tgacgcaaaa taccaccttg gcgcctacac
35    601 ggagagcagc gtccgcatcg tccgtgacga catgctgtgt gccgggaaca cccggaggga
661 ctcatgccag ggcgactccg gaggggccct ggtgtgcaag gtgaatggca cctggctgca
721 ggcggcgctg gtcagctggg gcgagggtg tgcccagccc aaccggcctg gcactacac
781 ccgtgtcacc tactacttgg actggatcca ccactatgtc ccaaaaaagc cgtgagtcag
841 gcctgggttg gccacctggg tcaactggag accaaccctt gctgtccaaa acaccactgc
40    901 ttcctacca ggtggcgact gccccccaca cttccctgc cccgtcctga gtgcccttc
961 ctgtcctaag cccctgctc tcttctgagc ccctccctt gtctgagga cccttcccc
1021 tcttgagccc cttccctgt cctaagcctg acgctgcac cgggccctcc ggccctcccc
1081 tgcccaggca gctggtggtg ggcgctaata ctctgagtg ctggacctca ttaaagtga
1141 tggaaatc.

```

## 45 SEQ ID NO: 262

Amino acid sequence of human TPSB2 encoded by the DNA sequence shown in SEQ ID NO: 261.

MLNLLLLALPVLASRAYAAPAPGQALQVRGIVGGQEAPRSKWPQVSLRVHGPYWMHFCG  
 5 GS LIHPQWVLTAAHCVGPDVKDLAALRVQLRQHLYYQDQLLPVSRIIVHPQFYTAQIGA  
 DIALLELEEPVKVSSHVHTVTLPPASETFPPGMPWCWVTGWGDVDNDERLPPFFPLKQVKV  
 PIMENHICDAKYHLGAYTGDDVRIVRDDMLCAGNTRRDSQQGDSGGPLVCKVNGTWLQAG  
 VVSWGEGCAQPNRPGIYTRVTTYLDWIHHYVPKKP

SEQ ID NO: 263

10 Amino acid sequence of human TPSB2, a soluble active secreted form derived from SEQ ID NO:262.

IVGGQEAPRSKWPQVSLRVHGPYWMHFCGGS LIHPQWVLTAAHCVGPDVKDLAALRVQL  
 REQHLYYQDQLLPVSRIIVHPQFYTAQIGADIALLELEEPVKVSSHVHTVTLPPASETFP  
 PGMPWCWVTGWGDVDNDERLPPFFPLKQVKVPIMENHICDAKYHLGAYTGDDVRIVRDDML  
 15 CAGNTRRDSQQGDSGGPLVCKVNGTWLQAGVVSWGEGCAQPNRPGIYTRVTTYLDWIHHY  
 VPKKP

SEQ ID NO: 264

gi|6857813|ref|NM\_010781.1| Mus musculus mast cell protease 6 (Mcpt6), mRNA

1 ctaagatgct gaagcggcgg ctgctgctgc tgtgggcact gtccctcctg gctagtctgg  
 20 61 tgtactcagc ccctcgccca gccaatcagc gaggtaggcat cgtgggagga catgaggcctt  
 121 ctgagagtaa gtggccctgg caggtagagcc tgagatttaa attaaactac tggatacatt  
 181 tctgaggagg ctctctcacc caccacagc gggtagctac tgcggcacac tgtgtgggac  
 241 cgcacatcaa aagccacag ctcttcggg tgcagcttcg tgagcagtat ctatactatg  
 301 gggaccagct cctctcttgg aaccggatcg tggtagcacc ccactattac acggccgagg  
 361 gtggggcaga cgtagccctg ctggagcttg aggtccctgt gaatgtctcc acccatatcc  
 25 421 accccatatt cctgccccct gcctcggaga ccttcccccc tgggacatcg tgctgggtga  
 481 caggctgggg cgacattgat aatgacgagc ctctcccacc tccttatcct ctgaagcaag  
 541 tgaagggtcc cattgtggaa aacagcctgt gtgaccggaa gtaccacact ggcctctaca  
 601 cgggagatga ttttccatt gtccatgatg gcatgctgtg tgctggaaat accaggagag  
 661 actcctgcca gccattctgt attggagatg acatctgatg agggtagacat ctacatttta  
 30 721 caattactgc tgagtgtga cctttccctc tcttgatgg aggcctcacc cttgatggga  
 781 aattaggtca gttgtccac ccagtgttag gtgtcatctc tggctctctg tgtgtgtcac  
 841 actcttcacc gctgaaccat ctctcgagtt ccttagtttt catttttaat gtcaaacata  
 901 acagactcat ccatcacaaa aataaaaggt gaatgtaaaa aaaa

SEQ ID NO: 265

35 Amino acid sequence of mouse TPSB2 encoded by the DNA sequence shown in SEQ ID NO: 264.

MLKRLLLLLWALSLLASLVYSAPRPANQVRGIVGGHEASESKWPQVSLRPFKLNWIHFC  
 GGS LIHPQWVLTAAHCVGPHIKSPQLFRVQLREQYLYGDQLLSLNRIIVHHPHYTAEAGG  
 40 ADVALLELEVPVNVSTHHPISLPPASETFPPGTSCWVTGWGDIDNDEPLPPPYPLKQVK  
 VPIVENS LCDRKYHTGLYTGDFFPIVHDGMLCAGNTRRDSQQPFCIGDDI

SEQ ID NO: 266

gi|9506886|ref|NM\_019180.1| Rattus norvegicus mast cell protease 6 (Mcpt6), mRNA



1 ggagagagga gccgagacag ccaagatgct gaagctgctg ctgctgctgg cactgtcccc  
 61 cctggctagt ctggtgcacg cgccccttg cccagtcaag cagcgagtgg gcattgtggg  
 121 aggcagagag gcttctgaaa gtaagtggcc ctggcaggtg agcctgagat ttaaattcag  
 181 cttctggatg ctttctgtg ggggtccct cattcaccca cagtgggtgc tcaactgcggc  
 5 241 aactgtgtg ggactgcaca tcaaaagccc agagctcttc cgtgtacagc ttcgtgagca  
 301 gtatctatac tatgcggacc agctactgac tgtgaaccgg accgttgtgc acccccacta  
 361 ctacacagtc gaggatgggg cagacattgc cctgctggag cttgagatcc ctgtgaatgt  
 421 ctccacccat atccacccca tatccctgcc ccctgcctcg gagaccttcc cctcggggac  
 481 ttcttgctgg gtaacaggct ggggcgacat tgatagtgc gagcctctcc tgccacctta  
 10 541 tcctctgaag caagtgaagg tcccattgt ggaacacagc ctgtgtgatc ggaagtacca  
 601 cactggcctc tacacaggag atgatgttcc cattgtccag gatggcatgc tgtgtgtcgg  
 661 aaataccagg agcgactcct gccagggaga ctcaaggggc ccactggtct gcaaaagtga  
 721 gggtaacctg ctgcaagcag gagtggtcag ctggggtgag ggctgcgcag aggccaatcg  
 781 tcctggcatt tacacccggg tgactacta cctggactgg attcaccgct atgtccctca  
 15 841 gcgttctctg gacccatcca gggtcagggg agaaccaggc acctgctgtc ttaactcac  
 901 tgcttctctg ccagatggaa ccctggcctt ctttgtactc tgtctccct gtctaccggg  
 961 tgtccctctg agcccccct ttgttccacc ttgagtccct cgccactcct gtcccctctg  
 1021 cctccacca caacacagct gcaactgtgc gctccctctt ttctgtggct cattaagta  
 1081 tgtgaaaatt ttgctcc

20 SEQ ID NO: 267

Amino acid sequence of rat TPSB2 encoded by the DNA sequence shown in SEQ ID NO: 266.

25 MLKLLLLLALSPLASLVHAAPCPVKQRVGIVGGREASESKWPQVSLRFKFSFWMHFCGG  
 SLIHPQWVLTAAHCVLHIKSPELFRVQLREQYLYADQLLTVNRTVVHPHYTVEDGAD  
 IALLELEIPVNVSTHHPISLPPASETFPSGTSCWVTGWGIDSDPELLPPYPLKQVKVP  
 IVENSLCDRKYHTGLYTGGDDVPVQDGMLCAGNTRSDSCQGDSSGGLVCKVKGTWLQAGV  
 VSWGEGCAEANRPGIYTRVTYYLDWIHRYVPQRS

SEQ ID NO: 268

30 gi|18491001|ref|NM\_003881.2| Homo sapiens WNT1 inducible signaling pathway protein 2 (WISP2), mRNA

1 tgtgtgtgtg tgtgtgtgag cgcgcgcgcg cgcgcgcgctg tgtactcgtg cgtgtgcctg  
 61 tgtgtgcctg ggagtgcact cacagctgcc ggaacataaa gactcacagg tccgcctccc  
 121 aggctcaaag ctggctctgc aggggacatg agaggcacac cgaagaccca cctcctggcc  
 181 ttctccctcc tctgcctcct ctcaaagggt cgtacccagc tgtgcccagc accatgtacc  
 35 241 tgcccctggc caccctcccg atgcccgctg ggagtacccc tgggtgtgga tggctgtggc  
 301 tgcctgcggg tatgtgcacg gcggctgggg gagccctgcg accaactcca cgtctgcgac  
 361 gccagccagg gcctggtctg ccagcccggg gcaggacccg gtggccgggg ggccctgtgc  
 421 ctcttggcag aggacgacag cagctgtgag gtgaacggcc gcctgtatcg ggaaggggag  
 481 accttccagc cccactgcag catccgctgc cgctgcgagg acggcggctt cacctgcgtg  
 40 541 ccgctgtgca gcgaggatgt gcggctgcc agctgggact gccccaccc caggagggtc  
 601 gaggtccttg gcaagtgtg ccctgagtgg gtgtgcggcc aaggaggggg actggggacc  
 661 cagccccttc cagcccaagg accccagttt tctggccttg tctcttccct gcccctgggt  
 721 gtcccctgcc cagaatggag caggcctgg ggaccctgct cgaccacctg tgggctgggc  
 781 atggccaccc ggggtgtcaa ccagaaccgc ttctgccgac tggagaccca gcgccgctg  
 45 841 tgctgtcca ggccctgcc accctccagg ggtcgcagtc caaaaacag tgccttctag  
 901 agccgggctg ggaatgggga cagggtgtcc accatcccca gctgggtggc ctgtgcctgg  
 961 gccctgggct gatggaagat ggtccgtgcc caggcccttg gctgcaggca acactttagc  
 1021 ttgggtccac catgcagaac accaatatta acacgctgcc tggctgtct ggatcccag  
 1081 gtatggcaga ggtgcaagac ctagtccctt tctcttaac tcaactgcta ggaggctggc  
 50 1141 caagggtgtcc aggttctct agccactcc ctgcctacac acacagccta tatcaaacat  
 1201 gcacacgggc gagctttctc tccgacttcc cctgggcaag agatgggaca agcagtcct  
 1261 taatattgag gctgcagcag gtgctgggct ggactggcca ttttctggg ggtaggatga

1321 agagaaggca cacagagatt ctggatctcc tgetgccttt tctggagttt gtaaaattgt  
 1381 tcctgaatac aagcctatgc gtgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa

## SEQ ID NO: 269

- 5 Amino acid sequence of human WISP2 encoded by the DNA sequence shown in SEQ ID NO: 268.

MRGTPKTHLLAFSLCLLSKVRTQLCPTPCTCPWPPPRCPGLGVPLVLDGCGCCRCVARRL  
 GEPDQLHVCDAQQGLVCQPGAGPGGRGALCLLAEDSSCEVNGRLYREGETFQPHCSIR  
 CRCEGGFTCVPLCSEDVRLPSWDCPHRRVEVLGKCCPEWVCGGGGLGTQPLPAQGPQ  
 10 FSLVSSLPFGVPCPEWSTAWGPCSTTCGLMATRVSNQNRFCRLETQRRCLSRPCPPS  
 RGRSPQNSAF

## SEQ ID NO: 270

Amino acid sequence of human WISP2, a soluble active secreted form derived from SEQ ID NO: 269.

15 RTQLCPTPCTCPWPPPRCPGLGVPLVLDGCGCCRCVARRLGEPCDQLHVCDAQQGLVCQPG  
 AGPGGRGALCLLAEDSSCEVNGRLYREGETFQPHCSIRCRCEGGFTCVPLCSEDVRLP  
 SWDCPHRRVEVLGKCCPEWVCGGGGLGTQPLPAQGPQFSGLVSSLPFGVPCPEWSTAW  
 GPCSTTCGLMATRVSNQNRFCRLETQRRCLSRPCPPSRGRSPQNSAF

## SEQ ID NO: 271

- 20 gi|8394540|ref|NM\_016873.1| Mus musculus WNT1 inducible signaling pathway protein 2 (Wisp2), mRNA

1 cccacgcgtc cgcgctcctg atctccagag gaccccgggc tgggacaggg gccttggcga  
 61 ggctgcagct gctgtggcag tagcttggga tggaggtctt tcttggctggg aactgaggag  
 121 ctgagaggct cctgtcaggc tcctgtccta aactcttggc acttgcggtg gcttgggctt  
 25 181 cacacactgt cagacacctt cttgggtggc tcctcggcct caggtttgaa gctggctcca  
 241 caagggaacac ggtgacatga ggggcaacc actgatccat cttctggcca ttctctcct  
 301 ctgcattctc tcaatggtgt attcccagct gtgcccagca ccctgtgcct gtccttggac  
 361 accacccacag tgcccaccgg ggtaccctt ggtgctggat ggctgtggct gctgtcgagt  
 421 gtgtgcacgg aggtctgggg agtcctgcga ccacctgcat gtctgcgacc ccagccaggg  
 481 cctggtttgt cagcctgggg caggccccag tggcctgggt gctgtgtgcc tcttcgaaga  
 541 ggatgacggg agctgtgagg tgaatggcgg caggtacctg gatggggaga cctttaaacc  
 601 caattgcagg gttttgtgcc gctgtgatga cgggtggttc acctgcctgc cgtgtgcag  
 661 tgaggatgtg cggctgcca gctgggactg cccacgcccc aggagaatac aggtgccagg  
 721 aagggtgctg cccgagtggg tgtgtgacca ggcagtgatg cagccggcaa tccagccctc  
 35 781 ctacagccaa ggacaccaac tttctgccct tgtcactcct gcactctccg atggccctg  
 841 tccaaactgg agcacagcct ggggccccctg ctcaaccacc tgtgggttgg gcatagccac  
 901 ccgagtatcc aaccagaacc gattctgcca actggagatc cagcgtcgcc tgtgtctgtc  
 961 cagaccctgc ctggcatcca ggagccacgg ctcatggaac agtgccctct agagccattg  
 1021 cggggatgtg gatacagggc ctgccattct cagcaaatgt ccctaggacc aggccttggg  
 40 1081 ctgatggtag atgccctct ccatgctctt ggctgcagtt aactgtcctg ggtgattca  
 1141 gtgtccagag cctctgagcg atccctgtct tgtctgaggt gggggaagca ggtgaccagc  
 1201 tccatttctc tggattctga cccaggtctc tgggttctcc tggctagtct ctcaaaactt  
 1261 cctgtatga aaaggacaac caaaaggacc tttaaagcta agctgtactg ggcaagcctg  
 1321 gccaccatgc tggggatagt gacagttaata ggtaccaggc agcagattgc ctgaaacatc  
 45 1381 cagggtccctt cttggacttc tatgtgcttg tcccaaagat tatgggtgac cttgtaagt  
 1441 tgcctttcct gatctgagaa caccctgccc ggctgggaag aattttctgg gaacatgaag  
 1501 agatggaatc acactattct taagagcgtt tgccaagtcc aggaacttga cctttgtatt  
 1561 tgtaaaaata cacatctctt aaatgctcac aaagcaagag gctccacact tctggcaggc

1621 cagggccttt ctcttcagca tgagagagac aaggaacagt agagtaccct cctctggagg  
 1681 actggcccgg tctggaataa acacccaaat caagtgtgga aaaaaaaaaa aaaa

SEQ ID NO: 272

- 5 Amino acid sequence of mouse WISP2 encoded by the DNA sequence shown in SEQ ID NO: 271.

MRGNPLIHLAISFLCILSMVYSQLCPAPCACPWTPPQCPGVPLVLDGCGCCRVCAARRL  
 GESCDHLHVCDPSQGLVCQPGAGPSGRGAVCLFEEDDGSCEVNGRRYLDGETFKPNCRVL  
 10 CRCDDGGFTCLPLCSEVRLPSWDCPRPRRIQVPGRCCEWVCDQAVMQPAIQPSSAQGH  
 QLSALVTPASADGPCPNWSTANGPCSTTCGLGIATRVSNNRQCQLEIQRRLLCLSRPCLA  
 SRSHGSWNSAF

SEQ ID NO: 273

gi|13928801|ref|NM\_031590.1| Rattus norvegicus WNT1 inducible signaling pathway protein 2 (Wisp2), mRNA

15 1 ctgcaaagat ctgacagacg cttctgatct ccagaggacc ctgggggtggg acaggggctt  
 61 ggcaaggctg cagcgtctggg cagtggcttg gaatggaggt ctttattact gggaactgag  
 121 gagctaagag gctcctgtca gcttgctcta aacccttagc acttggtggtg gcttgggctt  
 181 cacacactgt cagacacctt cgtggtggcc tccacggctc acctccagggt ttgaagctgg  
 241 ctccacaagg gacacgggtga catgaggggc agcccactga tccgtcttct ggccacttcc  
 20 301 ttctctctgccc ttctctcaat ggtgtgtgccc cagctgtgccc ggacaccctg cacctgtcct  
 361 tggacaccac cccagtgcctc acagggggta cccctggtgc tggatggctg tggtgtctgt  
 421 aaagtgtgtg cacggaggtc gacggagtc tgcgaacacc tgcattgtctg cgaaccacagc  
 481 cagggcctgg tttgtcagcc tggggcaggg cctggcggcc atggggctgt gtgtctcttg  
 541 gatgaggtg acgggtgactg tgaggtgaat ggccgcaggt acctggatgg agagaccttt  
 25 601 aaacccaatt gcaggggtcct gtgccgctgt gatgacgggt gcttcacctg cctcccgtg  
 661 tgcagtgagg atgtgacgct gccagctgg gactgcccac gcccacagag aatacagggtg  
 721 ccaggaaggt gctgccccga gtgggtatgt gaccagggag tgacaccggc gatccagcgc  
 781 tccgcggcgc aaggacacca actttctgccc cttgtcactc ctgctctgctc tgatgtctct  
 841 tggccaaatt ggagcacagc ctggggcccc tgctcaacca cctgtggggt gggcatagcc  
 30 901 acccgagtgt ccaaccagaa ccgattctgc caactggaga tccaacgcgc cctgtgtctg  
 961 cccagaccct gcctggcagc caggagccac agctcatgga acagtgtctt ctaaggccaa  
 1021 ctggggatgc ggatacaggg cctgccatcc tcagcaaattg acccttggac caggccctgg  
 1081 actgctggta gatgctcttc tccatgctct tggctgcagt taactgtcct gcttggattc  
 1141 actgtgtaga gccactgagc gatccctgct ctgtctgagg taggcggagc aggtgaccag  
 35 1201 ctccagttct ctgggtcagc ctggaattct ggggtctcct ggctcattcc tcaaaacatc  
 1261 cctgtacaaa aaggacaacc aaaaagacct ttaaacctag gctatactgg gcaaacctgg  
 1321 ccaccgtgct ggggataagg tcaatgtag gaccagacag cagattgcct gaaacttcca  
 1381 attcccttct tggacttctg tatgcttgta accaaagatg atgaatgaac tcgtaagtgt  
 1441 accttccctg acctgagaac accctgcctg ctcggaaggt attcaggggc agaattctct  
 40 1501 gtgaacatga agagatgaat cacactgtcc ttaagaaatt cctcaaagtc caggagcttg  
 1561 agctttgtat ttccaggaat gcacatctct taagcactcg caaacacagga aggtccaca  
 1621 cctctaacag ccagggcctt tctcttcagc atgagaaaga caagggacag cagagtactc  
 1681 tcgtctggag gactagtcta gcctagaata aacacccaaa tcaagcgtga aaaaaaaaaa  
 1741 a

- 45 SEQ ID NO: 274

Amino acid sequence of rat WISP2 encoded by the DNA sequence shown in SEQ ID NO: 273.

MRGSPLIRLLATSFLCLLSMVCAQLCRTPCTCPWTPPQCPGVPLVLDGCGCCCKVCARRL

TESCEHLHVCEPSQGLVCQPGAGPGGHGAVCLLDEDDGDCEVNGRRYLDGETFKPNCRVL  
 CRCDDGGFTCLPLCSEDVTLPSWDCPRPKRIQVPGKCCPEWVCDQGVTPAIQRSAAQGHQ  
 LSALVTPASADAPWPNWSTAWGPCSTTCGLGIATRVSNQNRFCQLEIQRRRLCLPRPCLAA  
 RSHSSWNSAF

5

1 WHAT IS CLAIMED IS:

2 1. A method for identifying an agent for treating an obese, diabetic or  
3 pre-diabetic individual, the method comprising the steps of:

4 (i) contacting an agent to a polypeptide encoded by a polynucleotide that  
5 hybridizes to a nucleic acid encoding SEQ ID NO: 213, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22,  
6 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68,  
7 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112,  
8 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150,  
9 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186,  
10 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 215, 217, 219, 221, 223, 225,  
11 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260,  
12 262, 263, 265, 267, 269, 270, 272 or 274 in 50% formamide, 5X SSC, and 1% SDS at 42°C  
13 followed by a wash in 0.2X SSC, and 0.1% SDS at 55°C; and

14 (ii) selecting an agent that modulates the expression or activity of the  
15 polypeptide or that binds to the polypeptide, thereby identifying an agent for treating an  
16 obese, diabetic or pre-diabetic individual.

1 2. The method of claim 1, wherein the polypeptide comprises an amino  
2 acid sequence at least 95% identical to SEQ ID NO: 213, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22,  
3 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46, 47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68,  
4 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112,  
5 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150,  
6 152, 154, 156, 158, 160, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 182, 184, 186,  
7 188, 190, 192, 194, 196, 197, 199, 201, 203, 205, 207, 209, 211, 215, 217, 219, 221, 223, 225,  
8 227, 229, 230, 232, 234, 236, 238, 240, 242, 243, 245, 247, 249, 251, 252, 254, 256, 258, 260,  
9 262, 263, 265, 267, 269, 270, 272, 274 or a protein domain thereof.

1 3. The method of claim 1, the method further comprising detecting  
2 whether the selected agent modulates weight and/or obesity.

1 4. The method of claim 1, the method further comprising detecting  
2 whether the selected agent modulates insulin sensitivity.

1                   5.     The method of claim 1, wherein step (ii) comprises selecting an agent  
2     that modulates expression of the polypeptide.

1                   6.     The method of claim 1, wherein step (ii) comprises selecting an agent  
2     that modulates the activity of the polypeptide.

1                   7.     The method of claim 1, wherein step (ii) comprises selecting an agent  
2     that specifically binds to the polypeptide.

1                   8.     The method of claim 1, wherein the polypeptide is expressed in a cell  
2     and the cell is contacted with the agent.

1                   9.     The method of claim 1, wherein the polypeptide comprises SEQ ID  
2     NO: 213, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 33, 35, 37, 39, 40, 42, 44, 46,  
3     47, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94,  
4     96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132,  
5     134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 161, 163, 165, 167, 169,  
6     171, 173, 175, 177, 179, 181, 182, 184, 186, 188, 190, 192, 194, 196, 197, 199, 201, 203, 205,  
7     207, 209, 211, 215, 217, 219, 221, 223, 225, 227, 229, 230, 232, 234, 236, 238, 240, 242, 243,  
8     245, 247, 249, 251, 252, 254, 256, 258, 260, 262, 263, 265, 267, 269, 270, 272 or 274.